D6.1 - Food Safety Risk Assessment User-driven Requirements & Use Cases

Co-funded by the Horizon 2020
Framework Programme of the European Union

DELIVERABLE NUMBER	D6.1
DELIVERABLE TITLE	Food Safety Risk Assessment User driven Requirements & Use Cases
RESPONSIBLE AUTHOR	Matthias Filter (BfR)

GRANT AGREEMENT N.	731001
PROJECT ACRONYM	AGINFRA PLUS
PROJECT FULL NAME	Accelerating user-driven e-infrastructure innovation in Food & Agriculture
STARTING DATE (DUR.)	01/01/2017 (36 months)
ENDING DATE	31/12/2019
PROJECT WEBSITE	plus.aginfra.eu
COORDINATOR	Nikos Manouselis
ADDRESS	110 Pentelis Str., Marousi GR15126, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Mrs. Georgia Tzenou

WORKPACKAGE N. | TITLE	WP6 | Food Safety Risk Assessment Community
WORKPACKAGE LEADER	German Federal Institute for Risk Assessment (BfR)
DELIVERABLE N. | TITLE	D6.1 | Food Safety Risk Assessment User driven Requirements & Use Cases
RESPONSIBLE AUTHOR	Matthias Filter (BfR)
REPLY TO	Matthias.filter@bfr.bund.de
DOCUMENT URL	http://www.plus.aginfra.eu/sites/default/files/deliverables/D6.1.pdf
DATE OF DELIVERY (CONTRACTUAL)	30 June 2017 (M6), 30 June 2018 (M18, Updated version)
DATE OF DELIVERY (SUBMITTED)	28 July 2017 (M7), 2 July 2018 (M18, Updated version)
VERSION | STATUS	V2.0 | Final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Matthias Filter (BfR), Taras Günther (BfR)
REVIEWERS	Pascal Neveu (INRA), Esther Dzale (INRA)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial draft	20/07/2017	Matthias Filter (BfR)
0.2	Minor changes (typos, Layout)	25/07/2017	Matthias Filter (BfR), Taras Günther (BfR)
0.3	Minor changes (Layout, Acronyms List)	28/07/2017	Matthias Filter (BfR), Taras Günther (BfR)
1.0	Review Processed, final version	28 July 2017	Taras Günther (BfR), Matthias Filter (BfR)
2.0	Updated version	02 July 2018	Taras Günther (BfR), Matthias Filter (BfR)

PARTICIPANTS		CONTACT
Agro-Know IKE (Agroknow, Greece)		Nikos Manouselis Email: nikosm@agroknow.com
Stichting Wageningen Research (DLO, The Netherlands)		Rob Lokers Email: rob.lokers@wur.nl
Institut National de la Recherché Agronomique (INRA, France)		Pascal Neveu Email: pascal.neveu@inra.fr
Bundesinstitut für Risikobewertung (BFR, Germany)		Matthias Filter Email: matthias.filter@bfr.bund.de
Consiglio Nazionale Delle Richerche (CNR, Italy)		Leonardo Candela Email: leonardo.candela@isti.cnr.it
University of Athens (UoA, Greece)		George Kakaletris Email:gkakas@di.uoa.gr
Stichting EGI (EGI.eu, The Netherlands)		Tiziana Ferrari Email: tiziana.ferrari@egi.eu
Pensoft Publishers Ltd (PENSOFT, Bulgaria)		Lyubomir Penev Email: penev@pensoft.net

ACRONYMS LIST

VRE	Virtual Research Environment
WP	Workpackages
QMRA	Quantitative Microbial Risk Assessment
CV	Controlled Vocabulary
GUI	Graphical User Interface
API	Application Programming Interface
RSS	Rich Site Summary
DOI	Digital Object Identifier
FSK-ML	Food Safety Knowledge Markup Language
FSKX files	Food Safety Knowledge Exchange File
OS	Operating System

EXECUTIVE SUMMARY

This report describes the findings from the analysis of requirements for AGINFRA+ services from the "Food Safety Risk Assessment" community use cases. Specific requirements have been delineated from a comprehensive use case definition and from workshops with external researchers of the food safety risk assessment community. As a result, this work identified numerous new features that could improve existing AGINFRA resources and make them even more relevant to the food safety risk assessment community. This deliverable also describes two specific use cases where existing and new AGINFRA+ services will be applied and tested.

TABLE OF CONTENTS

1 INTRODUCTION

Within the AGINFRA+ project two use cases of the Food Safety Risk Assessment domain were selected, that address two different research communities within the area.

In the first use case AGINFRA+ will support the "Emerging risk identification community". In this community, the development of services that facilitate the early identification of emerging issues and risks in the food (and feed) chain is the main objective. Such services are of paramount importance, as emerging risk identification is a key element in governmental consumer protection strategies that want to implement timely and effective preventive measures. Increased global trade is making food chains more complex, both in terms of geographical spread and the rapid distribution of goods. As a consequence, yesterday's emerging issue in one area may be tomorrow's crisis in a different place. Dealing with this complexity requires a high degree of scientific and technical expertise. The use of new data mining and data science solutions (digital technologies) are crucial to enable the identification of emerging food safety issues at an early stage. Therefore, this use case will support the emerging risks detection in the European Union through facilitating digital technologies and "big data" analytics. When doing so, the AGINFRA+ solutions will support primarily EFSA and all EU Member State authorities to share data, knowledge and methods in a rapid and effective manner.

The second use case is directed to support data-intensive applications powered by one of the FoodRisk-Labs software tools (https://foodrisklabs.bfr.bund.de/frl/). The primary focus is here the extension of FoodRisk-Labs' capabilities to support the sharing of mathematical models as well as to provide web-based services that allow the application of models in user-defined simulation scenarios. Through the envisaged AGINFRA+ resources risk assessors, modellers and risk managers could be enabled to exploit easy-to-maintain open food safety model repositories, which contain mathematical models from the area of predictive microbial modelling, quantitative microbial risk assessment (QMRA) as well as chemical risk assessments. As this use case also covers the need for performant model simulation infrastructure, the envisaged AGINFRA+ solutions could provide a significant benefit for many risk assessment and management authorities and for European research institutes as well. In principle, it would even be possible to extend these resources also for private companies, but this will not be explored within the AGINFRA+ project.

2 AGINFRA+ USE CASE DESCRIPTION

2.1 USE CASE 1: DETERMINATION AND METRICS OF EMERGING RISK - DEMETER

2.1.1 Main stakeholders and organizations

- Data scientist, Researcher, DEMETER project members
- National Food Safety authorities (e.g. BfR)
- EFSA
- EU-Member state representatives in EFSA's EREN network
- Regional Food Safety offices (German Federal States)

2.1.2 Use case description

The early identification of emerging risks in the food (and feed) chain is an important governmental and business activity carried out to protect the consumer with timely and effective preventive measures. The increasing complexity of food chains makes it increasingly difficult to oversee and assess all food safety risks in each particular chain. Identification, integration and analysis of information collected from public and non-public sources is therefore essential to support decision making of public and private sector stakeholders. New data mining and data science solutions can provide such solutions helping to identify emerging food safety issues at an early stage.

The objectives of the DEMETER community is therefore to create an open, transparent, modular web-based system, which will support processes established by EFSA and other food safety authorities for emerging risks identification. Specifically, it is desired to provide resources that allow sharing and executing emerging risk data retrieval and data mining pipelines as well as sharing data and knowledge in a broader sense. The envisaged system should have a user-management system in place and support the integration of non-public (e.g. governmental or supplier audits) information in a privacy protecting manner. A strong focus should be laid on a service-based system architecture that will provide the necessary flexibility for fast integration of new internal or external services (if desired by the community).

2.1.3 Personas of the Community

Risk Assessor / Modeller / Data Scientist:

Current frustrations:

Currently there is no real "Knowledge exchange portal" in place that goes beyond sharing text document. Data-analysis pipelines developed by individual EFSA member states or researchers are not shared within the community. "Knowledge" on emerging risks is currently often available only as text-based documents. Furthermore, there are only few visualisation options for information on emerging risks. Finally, existing tools (usually only available as desktop solution and partly based on commercial products) have issues with usability.

Risk Manager:

Currently there is no real "Knowledge exchange portal" in place that goes beyond sharing text document. Even a protected space for communication (e.g. a messaging board) is missing. In addition, there are currently no "easy-to-use" risk identification tools available for Risk Managers.

The following needs of the DEMETER community were identified:

2.1.4 Data and Relevant Semantics Needs

1. Easy access to open scientific literature and other free online information sources on the WWW
2. Easy access to social media data (Twitter, Facebook)
3. Easy access to RSS feeds from community information providers: e.g. MediSys
4. Access to ontologies to support automated knowledge generation and extraction
5. A service to develop and maintain controlled vocabularies (CV) / ontologies e.g. a combination of an ontology browser with a mind map. Desired functionalities:
    - Version control
    - Collaborative, parallel work by multiple authorized stakeholders of the community
    - Commenting function
    - Creation of URIs for all concepts within the CV or automatic mapping to corresponding terms in existing online ontologies (e.g. those in www.bioontology.org)

2.1.5 Data Analytics and Processing Needs

1. KNIME workflow execution including support for R and Python extensions
2. API access to integrated emerging risk identification services
3. Should “integrate” / “talk” with EFSA’s IT infrastructure “Zenodo”
4. Execution of a data mining workflow on a high-performance computing infrastructure (if necessary)
5. Docker functionalities for data mining workflows

2.1.6 Data Visualization and Publishing Needs

1. A service to streamline the publishing of models / data mining workflows to the scientific community (Publish in the sense of “creating a citable scientific publication” - ideally with a DOI)
2. A public service to search / filter for emerging risk identification models / workflows in a model / workflow catalogue (a very user-friendly GUI with search, sort, filter (and visualisation) options and the option for downloading one or multiple models / workflows (if the rights allow that))
3. Interactive online data and knowledge visualisation features
4. Support for community-driven curation processes for uploaded emerging risk identification models / workflowshttp://www.ebi.ac.uk/biomodels-main/
5. Visualization of Bayesian network models and predictions

2.1.7 Other Needs

1. User management system, i.e a component that allows managing the rights and accounts of community members, e.g. passwords, permissions, etc.
2. Data inventory / workspace, i.e. a storage space for electronic files for each community members. This component also needs to provide options for each user to share or publish files with others.
3. Tracing of documents, i.e. a component that allows creating a document history or so-called provenance reports.
4. Software inventory, i.e. a component that allows to share, document and work collaboratively on software projects or on small software code

5. Information resource inventory / knowledge base, i.e. a component that allows to share, create and work collaboratively on a community knowledge base, e.g. a content management system (CMS) with Wiki-like capabilities or a shared MindMap.
6. General messaging dashboard / chat function, i.e. a component that provides users with a project internal communication, notification and information exchange facility / mechanism, e.g. each user receives updates on news right after login or via email notification. The component must be customizable by the users.
7. Project management, planning and controlling features, i.e. a component that provides project managers and team members with necessary functionalities for project management and project work, e.g. a shared calendar, a shared address book, joint project planning features (reminder of deadlines, Gantt charts, tickets) etc.
8. Video conferencing / conference calls, i.e. a component that provide project managers and team members with necessary functionalities for conference calls or web meetings.
9. Collaborative online document editor, i.e. a component that allows collaborative editing of documents as e.g. in Google docs or Overleaf.
10. Online search and filter of documents, i.e. a component that allows searching and filtering documents inside the workspace respecting the access rights assigned to the documents and the users.
11. Educational resources, i.e. a component that helps the community to get used to the system's functionalities and user interface.
12. System administration and monitoring front end, i.e. component that allows system admins to monitor the system's performance, manage necessary update of integrated components and perform other administration tasks in a user-friendly graphical user interface (GUI).

2.2 USE CASE 2: KNOWLEDGE INTEGRATION PLATFORM - RAKIP

2.2.1 Main stakeholders and organizations

- BfR
- ANSES
- DTU
- EFSA
- Other EU Risk Assessment authorities
- global: FAO, other national Food Safety authorities

2.2.2 Use case description

Food safety as a global challenge requires efficient knowledge transfer between academia, business operators and governmental agencies. In Europe, a rich variety of useful models, software tools and databases for food safety risk assessment exists, but exchange of these kinds of information between different stakeholders is currently extremely difficult and time consuming. However, integration of mathematical models and modelling tools is vital to cope with the numerous existing and emerging food safety risk and challenges. Three European institutions specialized in food safety modelling and risk assessment (ANSES, BfR, DTU) currently collaborate in a joint effort to establish a Risk Assessment Modelling and Knowledge Integration Platform (hereinafter referred to as RAKIP) where the term "knowledge" specifically refers to data and models relevant for risk assessment tasks. The development of a RAKIP portal would improve transparency in data- or model-based risk assessments and facilitating the exchange of

knowledge between different software tools that are already available in each of the three institutions. Therefore it is desired to develop an open, community-driven Food Safety Model Exchange Portal, that allows to upload, review, execute, search and download risk assessment models and modules. An underlying requirement for such a resource is, that a harmonized data format (called Food Safety Knowledge Markup Language - FSK-ML) will be developed.

2.2.3 Personas of the Community

Risk Modeller / Data Scientist:

- A risk modeller wants to share his model with the scientific community by making it available to the public and useful. In order to accomplish this, he wants to upload his model to a community portal where it could be used online without much additional work to be needed, e.g. no need to re-implement his model into a different programming language.
- A risk modeller wants to create a new model. In order to accomplish it, the modeller wants to improve a model that is already available. For this, he would want to download a model from a public community portal, optimize the downloaded model to his needs and use/ share it afterwards.
- A risk modeller wants to create a new model. In order to accomplish it, the modeller wants to combine different models that are already available within the community portal.

Risk Assessor:

- A risk assessor wants to quickly generate a risk assessment for a given risk question raised by the Risk Manager. In order to accomplish this, the Risk Assessor wants to use existing curated models from the community platform that he "just" customize to the specific question of the Risk Manager. Usually, a Risk Assessor has not time for model re-implementation and therefore has to rely on ready-to-use models.

Risk Manager:

- A risk manager wants to take decisions for a given risk question. In order to accomplish this, the risk manager wants to look at the details of the risk assessment model including the simulation results generated by the Risk Assessor. Ideally, the risk assessment model could be executed online with the simulation settings provided by the Risk Assessor.

Lab Researcher / Food Microbiologist:

- A lab researcher/ food microbiologist wants to share experimental results, e.g. on the growth or survival of a microorganism under certain environmental or process conditions, with the scientific community. In order to accomplish this, the researcher/ food microbiologist wants to upload his data to the community portal without much extra work.

Current frustrations:

- Risk assessment models or experimental data are not available in a standardized data format
- No consensus within the risk assessment community on the metadata needed to annotate models or data
- No curated model repository infrastructure
- Models can be generated in different scripting languages - currently there is no platform available to bridge software language barriers (could be solved with the KNIME)

The following needs of the RAKIP community were identified:

2.2.4 Data and Relevant Semantics Needs

1. A service to develop and maintain controlled vocabularies / ontologies – see 2.1.4
2. Online resource to store / upload and create new models. The models should be provided in the FSK-ML format (i.e. as an FSKX file). The FSK-ML format specification is currently under development. Key feature of FSK-ML is, that this standard allows sharing script-based code that can be executed in the appropriate environment, e.g. KNIME.
3. A service that allow the user to visualize metadata or reconfigure FSKX model files from the model repository / workspace before execution / simulation

2.2.5 Data Analytics and Processing Needs

1. Need for KNIME workflow execution inside the VRE. These KNIME workflows will need to execute R based models. In the future other scripting language could be supported.
2. A service for model execution (simulation).
3. A service to use high-performance computing infrastructure in case computational expensive simulations need to be performed. Creating standardized provenance reports on the simulation process would be of extraordinary value.
4. API access to model simulation services.
5. Provide Docker functionalities inside the VRE to guarantee future re-execution of current models.
6. A service to connect to EFSA's data portal "EFSA Knowledge Junction".
7. A service that checks if a FSKX model is still executable (due to new software versions of OS, underlying scripting languages and third-party libraries this might not be true in the future)
8. Creation of URIs for each shared model.

2.2.6 Data Visualization and Publishing Needs

1. Interactive visualisation service for models (including model code and model metadata) provided as FSK-ML formatted files.
2. A service that read a FSK-ML formatted file with information on the QMRA model input and output parameters. The visualization service should allow users to modify input parameters (in a range defined by the model metadata) and get the simulation results represented in appropriate chart types instantaneously. An example how this could look like, is available here:
3. A service to combine model modules into new models
4. Establish a system for the curation of models
5. Interactive and user-friendly GUI of the model repository (including search, browse, sort, filtering functions), e.g. like

2.2.7 Other Needs

In addition to the list provided under 2.1.7 the following:

1. Data management policy is important: data storage and calculations should not be done on US server