

APPROVED: 6 July 2018

doi: 10.2903/j.efsa.2018.e16088

Application of data science in risk assessment and early warning

Bundesinstitut für Risikobewertung, Berlin, Germany,
Individual authors: Michał Jan Czyż, Matthias Filter and Anja Buschulte

Abstract

The currently applied approaches, procedures and tools used for the identification of emerging risks vary greatly among Member States of the EU. EFSA established a structured approach for emerging risk identification that mainly consists of systematically searching, collecting, collating and analysing information and data. In addition, EFSA concluded that new methodologies and tools are needed to facilitate efficient and transparent sharing of data, knowledge and methods in the field of emerging risk identification between Member States. As the result of an open call issued by EFSA, the 'Determination and metrics of emerging risks' (DEMETER) project was established in spring 2017 to support current and future procedures for identification of emerging risks. As the Bundesinstitut für Risikobewertung (BfR) hosting site is involved in the DEMETER project, as well as in several other software development activities in the area of quantitative microbiological risk assessment, the fellow had the opportunity to play an active role in the project work and development of the running DEMETER project. The training and close integration in the project team enabled the fellow to make significant contributions, e.g. with the creation of new open source data processing workflows and by contributing to the Emerging Risk Knowledge Exchange Platform (ERKEP) Framework Concept Note. Besides DEMETER, the fellow participated in other activities of the Unit for Food Technologies, Supply Chains and Food Defence, including testing and applying several BfR open source software tools which had been developed in previous projects and that are used in microbiological risk assessment (e.g. Predictive Microbial Modelling Lab (PMM-Lab)) or as automatic data retrieval systems (e.g. SiLeBAT NewsRadar) – see <https://foodrisklabs.bfr.bund.de>.

© 2018 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

Keywords: data science, emerging risks, early warning

Correspondence: eu-fora@efsa.europa.eu

Acknowledgements: This report is funded by EFSA as part of the EU-FORA programme.

Suggested citation: Bundesinstitut für Risikobewertung, Berlin, Germany, Czyż MJ, Filter M and Buschulte A, 2018. Application of data science in risk assessment and early warning. *EFSA Journal* 2018;16(S1):e16088, 7 pp. <https://doi.org/10.2903/j.efsa.2018.e16088>

ISSN: 1831-4732

© 2018 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

This is an open access article under the terms of the [Creative Commons Attribution-NoDerivs](https://creativecommons.org/licenses/by/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.



The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.



Table of contents

Abstract.....	1
1. Introduction.....	4
1.1. Bundesinstitut für Risikobewertung Unit 41	4
1.2. DEMETER project	4
2. Description of work programme	5
2.1. Aims.....	5
2.2. Activities/methods	5
3. Conclusions.....	6
References.....	6
Abbreviations	6

1. Introduction

The European Food Risk Assessment Fellowship Programme (EU-FORA) began in September 2017 (Bronzwaer et al., 2016). It was developed by members of the European Food Safety Authority (EFSA) (Fellowship Programme Committee) with additional support from representatives of Member State authorities. The main goal of the programme is to train a new generation of risk assessors. In the first edition, 15 early to mid-career scientists working in national authorities were placed for 1 year in the competent authority of another Member State. During this stay, each fellow was integrated into the work of the hosting site.

In addition to the work programme, all fellows were trained in microbiological and chemical risk assessment during a three-week introductory training at EFSA. Furthermore, three-one-week training courses covering different risk assessment subjects and risk communication were an integral part of the fellowship programme.

1.1. Bundesinstitut für Risikobewertung Unit 41

The project took place in the Department for Biological Safety (Dep. 4) of the German Federal Institute for Risk Assessment (Bundesinstitut für Risikobewertung (BfR)) and was supervised by the Department's Unit for Food Technologies, Supply Chains and Food Defence (Unit 41). The Department for Biological Safety deals with health risks to humans that may arise from microorganisms, the toxins formed by them and other microbial metabolites. This includes not only bacteria but also viruses, parasites and transmissible spongiform encephalopathy pathogens. The department is involved in establishing the cause of outbreaks of food-borne diseases and zoonoses. It has a number of national reference laboratories for the diagnosis and fine typing of pathogens, antibiotic resistance and the microbiological contamination of foods (a task anchored in food legislation). In this scope, Unit 41 deals with the identification and evaluation of hazards that may be present in food. Performing vulnerability assessments concerning these hazards and developing risk mitigation strategy concepts are related tasks of the unit. Unit 41 is furthermore involved in microbiological risk assessments and provides Germany's national expert for the EFSA Scientific Network on Microbiological Risk Assessment. Other key focuses are national and international research projects that aim to develop new data-driven infrastructure and knowledge supporting the efficient generation of risk assessments. In this context, several software tools have been developed to facilitate risk assessment. Examples are 'Predictive Microbial Modelling Lab' (PMM-Lab; aims to ease and standardise the statistical analysis of experimental microbial data and the development of predictive microbial models), 'FoodProcess-Lab' (for the application of predictive microbial models on food process chains) or 'Food Safety Model Repository' (openFSMR; a community-driven search engine for predictive microbial models). Other core activities are national and international research projects. Scientists from this unit were or are involved in many other projects, for instance: SiLeBAT, AniBioThreat, ZooGloW, FoodAuthent, AGINFRA+, DEMETER.

The fellowship programme was supervised by two senior research scientists from Unit 41: Dr Anja Buschulte, Veterinarian, Senior Research Scientist and Matthias Filter, Biochemist, Senior Research Scientist. Further support was given by other research scientists in the unit with many years of professional background in the field of either performing microbiological risk assessments or data deployment. Cross-unit activities provided the opportunity to get a closer insight into other risk assessment related issues.

Furthermore, the fellowship activities were integrated with the ongoing research project 'Determination and metrics of emerging risks (DEMETER)' that has received a grant from EFSA. This enabled the fellow to also learn from the other leading research institutes that are participating in this project, like Wageningen University, the Netherlands, the University of Newcastle upon Tyne, United Kingdom, and the National Food Chain Safety Office, Hungary.

1.2. DEMETER project

The general food law implemented in Regulation (EC) No 178/2002¹ obliges EFSA to establish monitoring procedures for the identification of emerging risks through searching, collecting, collating

¹ Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. OJ L 31, 1.2.2002, p. 1–24.

and analysing information and data. With regard to these requirements, EFSA has developed and implemented an approach to identify emerging risks. EFSA also recognised that the improvement of the coordination of resources, expertise and data across Europe and internationally, is of crucial importance. It concluded that new methodologies and tools were needed in order to facilitate efficient and transparent sharing of data, knowledge and methods in the field of emerging risk identification. As a result, an open call was issued by EFSA in which the DEMETER consortium was selected. DEMETER aims to support the current procedure for emerging risk identification by developing a collaborative platform to allow EU Member States and EFSA to share data, knowledge and methods for the identification of emerging food-related risks. One of the aims of this project was to develop a prototypic collaborative Emerging Risk Knowledge Exchange Platform (ERKEP).

The ERKEP was defined by the DEMETER project as 'a technical resource based on KNIME Server infrastructure hosted at BfR supporting the exchange of data analytics workflows for emerging risk identification'. The platform will be a technical solution to support the exchange of data analytics knowledge relevant for emerging issues and risk identification.

Besides this technical solution, EFSA asked the DEMETER project to develop a broader vision of what an 'ideal' knowledge exchange framework on emerging issues and risks should look like (ERKEP Framework). This vision includes specific concerns, e.g.:

- Which functionalities are desired?
- What should be shared?
- What side factors should be considered?

Several needs and functionalities were identified to be taken into account for the ERKEP Framework. The vision paper articulates the shared understanding about the values, mission, and approach that could guide EFSA's and DEMETER's efforts for the future development of resources for efficient knowledge exchange in regard to the emerging risk identification area.

2. Description of work programme

The fellow was associated with Unit 41 of BfR and had the chance to be involved in the DEMETER research project which started in spring 2017. The objectives and research proposed in this project were specifically designed to support current (and future) EFSA procedures for emerging risk identification by providing a set of integrated, open source solutions that would allow EFSA and EU Member State authorities to share data, knowledge and methods in a rapid and effective manner. As part of this work, the fellow had the opportunity to generate his own research results that could be published as a research paper or as an oral presentation or poster at scientific conferences.

2.1. Aims

The main goals of the fellowship project were:

- Learning best practice on data analysis principles (including transparency, validation and documentation).
- Getting insight into tasks of the department concerning microbiological risk assessment with emphasis on the application of scientific data for risk assessment.
- Getting familiar with software tools for data mining (KNIME) and predictive microbial modelling (e.g. PMM-Lab) and quantitative microbiological risk assessment (QMRA) (e.g. FSK-Lab).
- Analysing EFSA's emerging risk identification framework – current approaches and tools (e.g. MediSys, SiLeBAT NewsRadar).
- Generating automated data retrieval and monitoring pipelines using KNIME.
- Contributing to the design of DEMETER ERKEP and the ERKEP Framework vision.

2.2. Activities/methods

The fellow got close insights into the area of QMRA. He participated in BfR internal activities in this area and contributed to the ongoing development of software tools for the related community. Furthermore, the fellow was closely involved in the planning and conducting of activities in the DEMETER and other national and international research projects of the unit, e.g. AGINFRA+ and the national food project FoodAuthent. Among these activities were: collaboration in writing the DEMETER concept note, attendance at AGINFRA+ and DEMETER project meetings and contribution and

development of open source data analysis workflows and services for virtual research environments (VRE).

In addition to the individual training opportunities provided by Unit 41, the fellow had the opportunity to participate in the accompanying training programme by BfR for EU-FORA fellows. The training sessions and seminars were:

- Reviewing scientific literature systematically – an introduction.
- Risk assessment and risk management of genetically modified organisms.
- Presenting in English.
- Introduction to the BfR FoodRisk-Lab software tools.

As an extension to the main goals and primary agreements, a script book with introduction to the application of R scripting language in the area of risk modelling was written by the fellow. This tutorial will be made available as an open resource in an online open source repository and as an online course inside a VRE.

3. Conclusions

During AGINFRA+ and DEMETER project meetings in Wageningen (November 2017), the fellow presented sample workflows in the KNIME platform to automatically obtain, arrange and analyse data from social media and the European and Mediterranean Plant Protection Organization (EPPO) data services (EPPO, online).

In cooperation with other partners from the DEMETER project, an ERKEP framework vision paper was developed. Important parts of this document are the definition of a common vocabulary, the participants, and the current and future technical possibilities of both the ERKEP platform and framework.

The development of new workflows for data processing, as well as upgrading old workflows in collaboration with other Unit 41 members, resulted in providing open or better tools for emerging risk identification and assessment. Further tools are currently under development and will be deployed via the VRE or the BfR KNIME server infrastructure which serves as a basis for the DEMETER ERKEP prototype. Moreover, other developments on VREs are ongoing, including the creation of Wiki pages and improvements in the general VRE user experience.

Last, but not least, an introductory script book for using R language in risk assessment was developed. After review by members of Unit 41 and EU-FORA fellows it is available for the general public via the GitHub repository (Czyż, 2018). This open source project allows the broad community of R developers to share their expertise and knowledge in order to create a modern and comprehensive script book. As a result, risk assessors as well as others interested in modelling will have an opportunity to become familiar with R language which is one of the most appreciated tools in data analysis nowadays (Cass, 2017).

References

- Bronzwaer S, Le Gouierrec N and Koulouris S, 2016. Editorial: the European Food Risk Assessment Fellowship Programme (EU-FORA). *EFSA Journal* 2016;14(11):e14111, 2 pp. <https://doi.org/10.2903/j.efsa.2016.e14111>
- Cass S, 2017. The 2017 Top Programming Languages. *IEEE Spectrum*, 18 July. Available online: <https://spectrum.ieee.org/computing/software/the-2017-top-programming-languages>
- Czyż MJ, 2018. A Beginners Guide to R's Galaxy. Available online: <https://github.com/mczyzj/BeginnersGuideToGalaxy>
- EPPO (European and Mediterranean Plant Protection Organization), online. EPPO Data Services. Available online: <https://data.eppo.int/>

Abbreviations

AGINFRA+	Accelerating user-driven e-infrastructure innovation in food and agriculture
BfR	Bundesinstitut für Risikobewertung
DEMETER	Determination and Metrics of Emerging Risks
EPPO	European and Mediterranean Plant Protection Organization
ERKEP	Emerging Risk Knowledge Exchange Platform
EU-FORA	European Food Risk Assessment Fellowship Programme
FoodAuthent	Development of a system for collection, analysis, and utilisation of product data for authenticity in the food sector
FSK-Lab	Food Safety Knowledge Lab

openFSMR	Food Safety Model Repository
PMM-Lab	Predictive Microbial Modelling Lab
QMRA	quantitative microbiological risk assessment
SiLeBAT	Sicherstellung der Futter- und Lebensmittelwarenkette bei bio- und agro-terroristischen (BAT)-Schadenslagen
VRE	virtual research environment
ZooGloW	Zoonosen und Lebensmittelsicherheit entlang globaler Warenketten