

DEVELOPMENT OF A PILOT SERVICE FOR THE ELECTRONIC INFRASTRUCTURE OF METROFOOD-RI

*Presser K.*¹, *Zoani C.*², *Szymanek J.*¹, *Ocke M.*³, *Zappa G.*²

¹ Premotec GmbH, Winterthur, Switzerland, karl.presser@premotec.ch, jakub.szymanek@premotec.ch

² Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) - METROFOOD-RI Coordination Office, Roma, Italy, claudia.zoani@enea.it, giovanna.zappa@enea.it

³ National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands, marga.ocke@rivm.nl

Abstract – METROFOOD-RI is a research infrastructure to promote data and information on metrology in food and nutrition and is composed by a Physical infrastructure (P-RI) and an electronic infrastructure (e-RI). Concepts and the development of a pilot e-RI is presented in this work where several datasets from different countries are used and interrelated to integrate national e- resources into a European-wide e-RI providing new functionalities. The general functionalities will be described together with the challenges to combine already existing resources. Issues of the outcome will be discussed together with possible solution for the future.

Keywords: METROFOOD-RI, research infrastructure, system architecture, e-services

1. INTRODUCTION

The European Strategy Forum on Research Infrastructure (ESFRI), part of the European Commission, defines a research infrastructure to be facilities, resources and services used by a science community to conduct research and foster innovation [1], [2]. METROFOOD-RI is a long-term sustainable pan-European research infrastructure as defined by ESFRI and is currently undertaking its early phase while preparation and implementation are the next phases. This ‘early phase’ is funded by European Union’s Horizon 2020 research and innovation programme through the PRO-METROFOOD project (Progressing towards the Construction of METROFOOD-RI, G.A. 739568). METROFOOD-RI aims at providing high quality metrology services in food and nutrition, comprising an important cross-section of highly interdisciplinary and inter-connected fields throughout the food value chain. Three main goals are

formulated within METROFOOD-RI: First goal is to improve quality of data and strengthen cooperation and scientific knowledge by developing new metrological tools. The second goal is to provide a framework that will enable to carry out different research activities supporting data collection, measurement reliability, quality and safety, and traceability of food production, by realizing a network of plants, labs and experimental agro-farms. Third, to make available a new web platform for sharing and integrating information and data on metrological tools for food analysis and promoting the integration of existing database on food data. The presented work focuses on the third goal and describes concepts and implementations of a pilot e-RI for METROFOOD-RI.

2. BACKGROUND AND RELATED WORK

The e-facilities already available in the METROFOOD-RI consortium, together with the resources for hardware allocation have been inventoried and mapped. Based on this mapping, a classification of various resources was prepared and used to identify, define and describe the main e-services of METROFOOD-RI for the main user groups: laboratory users, research community, food business operators and consumers/citizens. E-resources are related to measurements, food production, food composition and food consumption. Different types of resources are available, such as: vocabularies, repositories, databases, software, models and e-learning platforms. As an example, databases supporting measurements are related to reference materials, official and reference methods, reference laboratories and proficiency testing providers. Databases related to food data concern: classification, food markers,

genetics, food profiling, food quality, food composition, contaminants in foods, microbiology, reference and threshold values. Models are available for food processing optimisation, meteorology and climate impact and environmental impact of food production.

Many of the existing database systems are lacking of web interfaces for users to access data and most of the systems are lacking of machine interfaces to exchange or export data by applications. Maintenance of the data is also often an issue. This situation makes it challenging for researchers, policy makers and the public to retrieve data. Most systems also use proprietary data formats which further increases the challenge to exchange data and a linkage of datasets for research purposes can only be done with additional efforts. In order to increase the value of these datasets, to better interface them with other existing e-networks and to fulfil the FAIR principle as defined by the European Commission (data should Findable, Accessible, Interoperable and Re-usable) [3], it is necessary to develop an appropriate electronic research infrastructure.

The architecture and concepts consider the current landscape of e-resources and e-infrastructures in the health and food domain. In particular, the e-resources made available by ELIXIR, MIRRI, BBMRI-ERIC and INSTRUMENT. Specifically concerning food composition, the reference networks at European and International level are EuroFIR AISBL and FAO/INFOODS respectively. In particular, EuroFIR implemented a high-level web application (FoodExplorer) where food composition datasets from most European countries are included, and data can be searched and compared [4].

To exchange data, a common data format or an explanation of what a data format means (ontology) is necessary. There are some existing data formats available from EuroFIR AISBL [5], the FP7 project TDS-Exposure [6] or EFSA [7]. These data formats were used as basis and extended where necessary with METROFOOD-RI information.

3. CONCEPTS and APPROACH

The METROFOOD e-RI will provide access to databases and tools with a new web platform (through its website www.metrofood.eu, once properly updated). In order to test the actual capability to deliver e-services, a pilot e-service was implemented in the frame of PRO-METROFOOD, providing an inventory of multiple datasets and a pilot web application to access datasets and make data browsable, searchable, filterable and comparable. An additional goal was to enable an existing application to retrieve and use data over machine interfaces. This use case should demonstrate the usability of the e-RI for data operation tools which can get single data records or complete datasets and can then perform some operations on them. The pilot service is focused to two classes of food: molluscs and rice.

In a first step to define an e-RI, a data management plan (DMP) was written. The DMP addresses what data will be generated, collected, handled and preserved following the FAIR principle. It also defines general access and restriction policies of the integrated datasets and what needs to be considered for data exchange.

Based on the DMP, the overall architecture of the pilot e-RI is designed as a distributed infrastructure as data providers are spread across Europe. A schematic overview of the architecture is given in Figure 1.

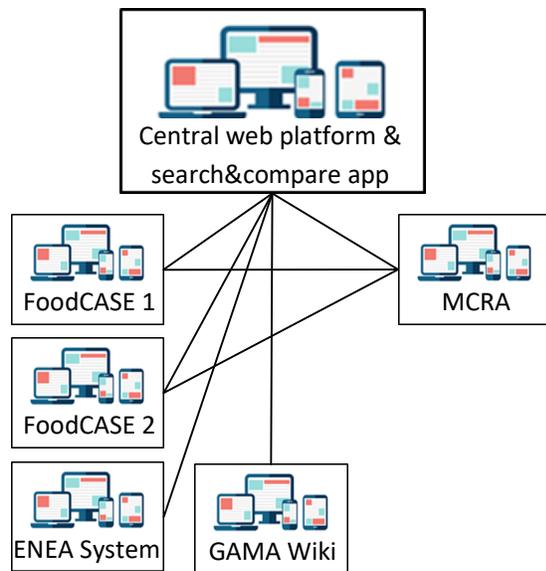


Figure 1. Schematic architecture of the pilot e-RI.

The central web platform consists of a landing page, an inventory and a search and compare application working on more datasets in parallel.

The central platform will be an inventory of data providers, datasets and tools available in METROFOOD-RI.

The search and compare application applies searches against providers' databases using RESTful JSON application programming interfaces (APIs). For the pilot implementation, two FoodCASE systems are used containing different datasets. FoodCASE is a food data management system that can manage and generate datasets for food composition, food consumption and Total Diet Study (TDS) data [8]. FoodCASE was selected to manage sampling analysis data as it already had some web services available for web applications. The architecture facilitates access policy and intellectual property rights (IPR) by keeping data on data owners' side. Where providers restrict access to their data, credentials will be used for authentication and authorisation to providers' system. RESTful APIs can also be used to exchange data with data operation tools or other RIs.

MCRA is an online tool to calculate exposure assessment by combining food consumption data and data on the levels of chemical compounds in foods [9]. MCRA was chosen as the data operation tool as it is well known in Europe and cooperation with EFSA exists. MCRA has interfaces to the two FoodCASE systems and will be listed as a data operation tool on the central web platform.

In addition, a web application developed by ENEA was chosen that is able to make an inventory of available resources on reference materials, proficiency testing schemes, reference methods and threshold limits for food analysis [10]. The ENEA system will be listed in the dataset inventory on the platform.

There is also an online wiki, called GAMA wiki, which was created in former projects and contain comprehensive information for macronutrients, vitamins, minerals and trace elements addressing all aspects of analytical procedures that were collected from international standards and scientific literature [11]. The wiki is interesting to integrate because it is a different form of database where whole pages of information are stored and data is not stored separately in relational tables. Some more datasets will also be listed on the platform which are not presented in Figure 1.

Each of these data provider systems can have own web applications to allow users to search data. The added value of the central web platform is that searches can be performed over all datasets and tools can be used to perform data operations.

For data exchange within and outside METROFOOD-RI, a data model was defined. The exchange format can easily be derived from the data model and JSON will be used as default format. Experiences in former projects shows that XML as exchange format is going to be large in size and is difficult to handle [12], [13]. As JSON is a general-purpose file format, it can be particularly useful for third party tools, which will be able to access data from METROFOOD-RI in a unified way.

APIs were defined to standardise the interfaces of the different systems and these APIs must be implemented by all data exchanging systems.

Each data provider needs to follow the exchange format and APIs to be compliant to the METROFOOD-RI and get involved in the e-RI.

4. RESULTS AND DISCUSSION

A first version of the pilot web platform was implemented and will be finalised and made accessible during the PRO-METROFOOD project. The pilot web platform serve as demonstrator how the e-RI can be implemented and as a discussion basis for further implementations. The pilot is also a playground to gain experiences with the distributed e-RI of METROFOOD-RI where different scenarios and configuration can be tested and evaluated.

The central web platform was implemented using WordPress. WordPress is a content management system (CMS) and was chosen as an open-source platform with a large community and many add-ons. As the platform also contains static content, a CMS is appropriate so that non-IT persons can modify content. The central data search and compare application is integrated into WordPress and is from IT point of view a new approach to combine a CMS and a bigger web application into one tool.

During the implementation, several issues needed to be investigated and solved and some of them are still in evaluation phase.

One of the issues is performance of the search and compare application. If the application needs to query, for instance, 45 database systems and each of the system needs two seconds to respond, then a waiting time of 90 seconds will result if the querying works sequentially. Data caching is an option but the whole research infrastructure will have a big amount of data and updating problems must be solved. The preferred way is to improve performance on each of the data provider systems, execute queries in parallel and make the search and compare application showing results as they are received. The application also needs to handle if some systems are not

responding or have a delay because of traffic. More experience will be needed and will show which approaches are feasible and which not.

Another issue was detected with access restrictions on certain data. Although the general rule in METROFOOD-RI is to make data open access, there is some data that with restricted access. The central web platform will have a user administration so that only registered users have access to the inventory and the search and compare application. On the other hand, each of the data provider systems have their own user administration and therefore own authentication and authorisation approaches. A first solution to forward user credentials to FoodCASE was found and tested. But with integration of more systems, a central authentication and authorisation mechanism will be needed.

A further issue was detected in the maintenance of the individual database systems. The distributed approach has advantages of cost sharing and control of data ownership but multiple players are responsible for the e-RI. This leads to more administration work and a surveillance system is needed to monitor the whole e-RI to guarantee a certain quality level.

ACKNOWLEDGMENT

PRO-METROFOOD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 739568. We also would like to thank EuroFIR for their cooperation and support, in particular on the GAMA wiki. We also thank all involved project partners for the cooperation.

REFERENCES

- [1] European Commission, https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=about, accessed 26.09.2017.
- [2] European Strategy Forum on Research Infrastructure, Strategy Report on Research Infrastructures, Roadmap 2010, 2011.
- [3] European Commission, Directorate-General for Research & Innovation, Guidelines on FAIR Data

Management in Horizon 2020, 2016, Version 3.0, 26.07.2016.

- [4] H. Pakkala, B. Koroušić Seljak, A. Kadvan, S. Bell, P. Finglas, The EuroFIR Food Platform: Further integration, refinement and exploitation for its long-term self-sustainability, Report on the development of FoodEXplorer (eSearch2) and future maintenance and operational requirements, EuroFIR NEXUS report, 2013.
- [5] W. Becker, A. Møller, J. Ireland, M. Roe, I. Unwin, H. Pakkala, Proposal for structure and detail of a EuroFIR Standard on food composition data, II. Technical Annex, EuroFIR TECHNICAL REPORT 1.2, 2008
- [6] K. Presser et al., Software Requirement Specification of FoodCASE-Risk, TDS-Exposure project, grant no. 289108, Deliverable 6.2, 2012
- [7] European Food Safety Authority (EFSA), Standard Sample Description ver. 2.0, EFSA Journal, 2013, 11(10):3424, 114 pp.
- [8] K. Presser, FoodCASE: A System to Manage Food Composition, Consumption and TDS Data, Food Chemistry, 2016 DOI: <https://doi.org/10.1016/j.foodchem.2016.09.124>
- [9] P. Boon, H. van der Voet, W. Boer, J. Kruisselbrink, M. van Lenthe J. van Klaveren, Monte Carlo Risk Assessment (MCRA) software: maintenance and management 2016, RIVM Report 2017-0014, 2017.
- [10] G. Zappa, C. Zoani, Networking Tools for Standardization, XXI IMEKO World Congress, 2015, Prague, Czech Republic.
- [11] I. Castanheira, M. Saraiva, A. Rego, V. Ollilainen, EuroFIR guidelines for assessment of methods of analysis: GAMA, Food Chemistry, 2016, Volume 193, pp. 82-89
- [12] A. Møller, T. Christensen, EuroFIR Web Services - EuroFIR Food Data Transport Package, Version 1.4, EuroFIR Nexus Technical Report D2.1 rev 1, 2012
- [13] H. Pakkala et al, EuroFIR Web Services – Specification of Request-Response Message Exchange Patterns, Version 1.2, EuroFIR Technical Report D1.8.29, 2