EU-FORA SERIES 5

Developing a framework for open and FAIR data management practices for next generation risk- and benefit assessment of fish and seafood

Javier Pineda-Pampliega, Annette Bernhard, Rita Hannisdal, Robin Ørnsrud, Gro Haarklou Mathisen, Gisle Solstad and Josef Daniel Rasinger

Institute of Marine Research (IMR), Norwegian Scientific Committee for Food and Environment (VKM), Norway

Abstract

Risk and risk–benefit assessments of food are complex exercises, in which access to and use of several disconnected individual stand-alone databases is required to obtain hazard and exposure information. Data obtained from such databases ideally should be in line with the FAIR principles, i.e. the data must be Findable, Accessible, Interoperable and Reusable. However, often cases are encountered when one or more of these principles are not followed. In this project, we set out to assess if existing commonly used databases in risk assessment are in line with the FAIR principles. We also investigated how access, interoperability and reusability of data could be improved. We used the OpenFoodTox and the Seafood database as examples and showed how commonly used freely available open-source tools and repositories can be implemented in the data extraction process of risk assessments to increase data reusability and crosstalk across different databases.

Correspondence: eu-fora@efsa.europa.eu

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

Table of contents

1. Introduction

2. Description of work programme

2.1. Aims

This project assessed how to apply FAIR data principles (Wilkinson et al., 2016) in risk and risk–benefit assessments of food. Focusing on key databases recently used in a risk–benefit assessment of fish and seafood in the Norwegian diet (VKM et al., 2022a), the OpenFoodTox and the Seafood database, we aimed to demonstrate how open-source software tools can be used to make data stored in publicly available repositories more Accessible, Interoperable and Reusable.

2.2. Methods

Using the programming language R (R Core Team, 2021) and data obtained from the ESFA OpenFoodTox Tool (Dorne et al., 2020; Kovarich et al., 2022) and from the Institute of Marine Research (IMR) Seafood database (Institute of Marine Research, 2022), we assessed if programmatical optimisation of access and the creation of a web-tool for selection and merging of subsets of the stored data improved accessibility, interoperability, and reusability of the data. In this section, a brief description of both data and tools used is provided.

2.2.1. R

The programming language and environment R has been designed for the statistical analysis of data and the creation of graphics (R Core Team, 2021). Over the past years, R has increasingly gained interest in the scientific research community (Hackenberger, 2020) as it is effective for data handling and includes many tools for basic and advanced data analysis (R Core Team, 2021). R is a well-developed non-static language, which means that its base features can easily be extended via packages that can provide new functions and functionalities for different data science challenges, including bioinformatics and data mining (Giorgi et al., 2022). In addition to this, R is supported by a big open-source community actively using this language and continuously adding new functionalities. R is licensed under the terms of the Free Software Foundation's GNU General Public License in source code form (R Core Team, 2021). To facilitate the programming with R, we used RStudio, an integrated development environment for R (RStudio Team, 2022).

2.2.2. Shiny

As commented above, R can be expanded through packages, being one commonly used 'Shiny' (Chang et al., 2021). This package was designed with the idea of creating interactive web applications which use R in the backend. While the creator of a web-based Shiny-tool does need to know R, the end user of the web application created with Shiny does not need to have any knowledge of R. In addition to local installations of R and Shiny, Shiny web app also can be stored on a server, which users can access through their web browser. In both cases, the appearance and functionalities of the applications are the same, and the underlying R code can be shared freely.

2.2.3. Git and GitHub

Git is a version control system designed to allow different users to work on the same programming project ensuring the traceability of progress and changes in the project. One of the most widely used providers of internet hosting for software development and version control using Git is GitHub (Microsoft, 2022). GitHub implements Git and offers a free version, in which users can host different smaller projects and scripts providing an easy way to share codes created in R and other programming languages on the web. The scripts generated during this project will be hosted and accessible on GitHub in this repository: https://github.com/J-Pineda-Pampliega/EU_FORA_Project (Pineda-Pampliega, 2022).

2.2.4. Zenodo

Under the European OpenAIRE program, and with the idea of championing the sharing of scientific data, the Zenodo (European Organization for Nuclear Research, and OpenAIRE, 2013) open repository

was developed and operated by CERN (European Council for Nuclear Research). This open-source repository was developed for scientific data in a broad way, allowing to deposit not only research papers, but also data sets, software, reports, supplementary data and any other research-related digital artefacts. Submissions to Zenodo obtain a persistent digital object identifier (DOI), which facilitates the citation of the stored items and allows the sharing of data prior to their publication in peer-reviewed journal.

For a speedy exchange of evidence and supporting materials which could be used in food and feed safety risk assessments, EFSA has created a curated open repository called the 'Knowledge Junction' within Zenodo. In addition to EFSA, several other institutions use Knowledge Junctions to share different data related to food security. For example, The Norwegian Scientific Committee for Food and Environment (VKM), which is part of this project, uses this Zenodo repository to upload finished reports (i.e. risk assessment and risk–benefit assessment) and supplementary materials of interest (i.e. literature searches, datasets, codes, etc.). To date, for VKM, the most recent example of the use of Zenodo is the opinion on the 'Risk-Benefit Assessment of Sunscreen' (VKM et al., 2022b). For this opinion, the fellow Javier Pineda-Pampliega contributed to the preparation of the public sharing of the report's supplementary material including datasets and R codes currently hosted on the VKM Knowledge Junction (Norwegian Scientific Committee for Food and Environment (VKM), 2022).

Zenodo recently implemented the possibility to import GitHub workspaces; it now is possible to host completed GitHub projects also on Zenodo. This offers the advantage of obtaining a DOI for one's code, which simplifies the traceability and proper citation of code used to create the results.

2.2.5. OpenFoodTox

The EFSA's Chemical Hazards Database (OpenFoodTox) (Dorne et al., 2020; Kovarich et al., 2022) is a structured database summarising the outcomes of hazard identification and characterisation for human and animal health and for the environment. It includes all regulated products and contaminants and provides open-source data for the (1) substance characterisation, (2) links to EFSA outputs, and the values of (3) reference points, (4) reference values and (5) genotoxicity. This database has become an essential tool for risk assessors and has provided the basis for the development and implementation of new approach methodologies (NAMs) in food and feed safety research. OpenFoodTox is hosted both on the EFSA webpage (as an interactive web tool) and on Zenodo in the EFSA Knowledge Junction.

2.2.6. Seafood database

The Institute of Marine Research in Norway routinely collects samples of key marine species for national and international monitoring programs. Their ISO 17025 accredited laboratories perform analyses of contaminants and nutrients using state-of-the-art methods. All the data generated, comprising multiple data points for over 25,000 individuals collected over a period of up to 15 years are aggregated in a large in-house database. This database can be accessed freely through the online Seafood database portal (Institute of Marine Research, 2022), where the user can select between fish, shellfish and seaweed divided by wild or farmed, and even prepared products, which can be found in Norwegian supermarkets. The database holds data of both Nutrients (separated into five categories: Amino acids, Fatty acids, Macro nutrients, Minerals and Trace elements and Vitamins) and Contaminants (separated into four categories: 'Drug residues', 'Heavy metals', 'Organic pollutants' and 'Other undesirable substances').

2.3. Activities

With the aim to investigate the application of FAIR data principles in risk–benefit assessment of seafood, it was essential to evaluate opportunities and limitations in the OpenFoodTox and the Seafood database. Once evaluated, we developed publicly available R and Shiny code, which attempts to address potential limitations found and to add new functionalities for sub-setting and improved crosstalk between hazard and occurrence data repositories.

2.3.1. Evaluation and actions on the OpenFoodTox

The OpenFoodTox database can be used in two different ways. The first (1) option is through the EFSA-hosted web application. The EFSA-hosted web application of the OpenFoodTox tool presents a classical interface, where different compounds can be searched by name. When searching, selected

substances appear in five different categories of results: Substance characterisation, EFSA outputs, Reference points, Reference values and Genotoxicity. The resulting output represents the main limitation, as each category only can be downloaded individually (either in pdf, csv or xlsx format). In other words, after a search, the users need to download five different files and manually merge the data.

The second (2) option to access data is to download the entire OpenFoodTox database in xlsx format (Microsoft Excel Open XML Spreadsheet) from Zenodo. The data comprises five individual spreadsheets providing data on (1) Substance characterisation, (2) EFSA outputs, (3) Reference points, (4) Reference values, and (5) Genotoxicity results. There is another "complete" spreadsheet which is a combination of the five spreadsheets commented above (each one in a different tab) in addition to a Dictionary spreadsheet (Dorne et al., 2020). This makes data Interoperable. However, as was described in the example above, to work with subsets of data spreading across the different spreadsheets, data aggregation and merging again must be performed manually using additional software for tabular data files. The most common among these tools is Excel, which is part of the commercial Microsoft Office Suite, but other free alternatives such as OpenOffice, LibreOffice or online tools such as Google Drive Sheets also can be used. In any case, for merging the large individual datasets, the user needs to be proficient in the terminology of terms and use of spreadsheet tools for efficient filtering, merging and sub-setting of the data in the desired format.

To evaluate potential complementary solutions to access, subset, and merge data stored in the EFSA OpenFoodTox database on Zenodo, in the present project using R (vers. 4.1.2) running in RStudio (vers. 2022.2.3.492), functions (i.e. pieces of code which work together for a common purpose) were written using R markdown, being characterised by the following features:

- Data is downloaded directly from the OpenFoodTox URL to eliminate the need for the user to search for and/or download the data in Excel.
- The possibility to search for up to 15 elements at the same time, with an implemented control of any repeated entry values. In the case of repetition, the repeated value is indicated, but not considered in the search.
- If a search is entered for a general term and several compounds appear in the database, an indication for the number of the different compounds is provided. For example, the search "lead" returns "4" results, because the components identified in the database are: "Lead", "Lead (II)", "Lead sulphate" and "Tetraethyl lead".
- To increase the (computational) reusability of the data in automated analysis pipelines, the information is downloaded in a plain text file (txt). This is a standard format of plain text that can be open in many different software tools. However, also the possibility to download data in csv (comma-separated values) is provided.
- To increase traceability information on the OpenFoodTox database version and the date and time when the file was created are automatically appended to the name of the downloaded file.

After the creation of the R script, to increase the number of potential users of this tool, we assessed if an additional approach that does not require knowledge and use of R could be developed. For this, the creation of a web-based application using Shiny was attempted. The use of Shiny opens the possibility to access and subset OpenFoodTox data using an internet browser only and also allows for the implementation of additional functions into our R code. That is, in addition to the characteristics of the function described above, the Shiny application developed in this project (Figure 1) has the following extra functions:

- Increased traceability: an indication of which version of the OpenFoodTox database used has been included. At the time of writing this report, the 5th iteration of the OpenFoodTox was released (and published on Zenodo on 16 June 2022).
- Implementation of interactive tables, allowing to filter results in real-time.
- Initially, tables will show all columns in the dataset but tools for sub-setting and selection of individual columns to be retained are provided. This functionality makes it easier to take snapshots only of the columns of interest for further uses.
- With one of the objectives of this project being to facilitate the interaction and crosstalk between databases of interest to risk assessors, the option to add links to PubChem for each selected compound was implemented. PubChem is a database of chemical molecules and their activities, maintained by the National Centre for Biotechnology Information (NCBI) of the United States.

2.3.2. Evaluation and actions on the Seafood database

The Seafood database contains information collected over a period of up to 15 years, with different data points for over 25,000 individual samples. This represents a comprehensive data repository of nutrients and contaminants in fish and seafood comprising more than 700,000 records. Due to the experience gained in the previous work with the OpenFoodTox tool, we directly designed a web application using Shiny to work with the Seafood database. As with OpenFoodTox, the first step was to evaluate the potential limitations and challenges of the existing system to access the database, which for the general public currently occurs via a web interface (https://sjomatdata.hi.no/). Having gotten access to the data underlying the web-based tool hosted at the IMR, in the present project, we assessed alternative solutions by addressing issues of the current web application using R and Shiny (Figure 2). We also set out to include additional functions potentially of use to risk assessors. The Seafood database Shiny web application is characterised by the following features:

- The publicly available web interface of the Seafood database is not version controlled. Furthermore, it is not updated with a defined periodicity, as it depends on data from different projects which are made available at different times throughout the year. This could be a challenge for the traceability of results and repeatability of analysis. As an attempted solution, we suggested for the database to be version controlled and to be updated at defined intervals only, e.g. annually. In addition, we implemented code to show a message highlighting the date when the database was last updated (Figure 3A). In a new version of our code, we also will include a button in the Shiny app to select which version of data the user wants to retrieve (i.e. to select the data regarding the day of the update).

- One common situation users of the Seafood database often encounter is the interest in the comparison of the presence of different compounds in different species or products. In the current web interface of the Seafood database, to check all the substances evaluated, it is only possible to select species or products one by one. In addition, to compare the concentration of different substances between species or products, the maximum number of substances is 10 by search. This makes it difficult to prepare a subset of desired data for further comparisons downstream. As a solution, in the prepared Shiny-based application, the user can select up to 15 species or products simultaneously, with information on all nutrients or contaminants. In addition, if the user is interested in only a particular set of compounds, up to 15 nutrients and another 15 contaminants can be selected.

- The R of FAIR means reusability of the data. This implies that for performing additional data analyses not yet envisaged by the data providers, users of a database should be able to access data presented in a non-aggregated way. Currently, the Seafood database does not provide this option; the results of searches are presented as numerical summaries (with sample size, mean, minimum and maximum values for each parameter). This makes it difficult to reuse this data in new evaluations. In the present project, at the IMR access to all data contained in the Seafood database was provided and two tables are presented in the Shiny application developed: one with a summary of the data (as in the IMR web interface), and another table with the non-aggregated data (Figure 3B).

- Continuing with reusability; in addition to access to non-aggregated data, also the format in which data can be downloaded by the user is important to consider. The Seafood database allows downloading in 'Portable Document Format' (pdf) format only. This format is widely used to present documents which include text and images and has the advantage of being immutable, i.e. independent of application software, hardware and operating systems documents are displayed always in the same way. However, this characteristic is a weakness for sharing data intended to be used in downstream analyses. For this, the data needs to be Reusable and Interoperable. The newly developed Shiny application allows for the download of selected data in txt or csv, being the most typical format to share data which could be used for further analysis. Both data from the summary table and the non-aggregated data can be downloaded in the desired formats. In addition, to ensure traceability when files are downloaded, the name consists of the date and the time of the creation and also incorporated the version of the database (the date of the latest update of the data; Figure 3C).

In addition to addressing specific limitations of the Seafood database listed above, we added extra functionalities in the Shiny code that we considered could be useful for the users:

- An increase in the number of inputs could entail an increase in the number of mistakes due to a repetition of terms in a search. To avoid this, our application indicates if a value is repeated but, even more important, the repeated value is not considered for the search, showing the same results as if the value was introduced only once (Figure 3A).
- Despite quality control measures in place, databases may contain erroneous entries (e.g. the inclusion of text in numeric rows and vice versa, or empty values). The Shiny application developed here includes a filter to flag and eliminate any rows which potentially contain mistakes. In addition, an indication of the number of eliminated entries is provided (Figure 3A).
- During the quantification of substances, it is possible that values are below the limit of quantification (LOQ) of a specific analysis. The LOQ is the lowest concentration of an analyte that can be quantified with a given certainty. In the Seafood database, values for contaminants below the LOQ are routinely reported using 'Upper bound' summation where the LOQ is used as if it were the actual concentration measured. This may result in many data points of the same value and such data sets are referred to as 'Left censored'. The web interface of the Seafood database indicates which values are below the LOQ, and also lists the numerical value of the respective LOQ for each method and compound in question. The Shiny application developed here, in addition, allows for further modification of the data and the possibility to calculate 'Lower bound' (substitute values < LOQ by 0) or 'Upper bound' (substitute values < LOQ with the actual LOQ) (Figure 3D).
- To increase data interoperability and crosstalk between different databases, common unique identifiers must be found. In our opinion, codes of the chemical substances in question provide a good option. Different unique identifiers do exist including InChI (International Chemical Identifier) or SMILES (Simplified Molecular Input Line Entry System), which both are included in the Seafood database. In addition to these, the 'paramCode' was added in the Shiny App, which is suggested by EFSA to be used when reporting on different substances in food and feed (EFSA, 2019).

One general challenge we found when working with the Seafood database is that its web interface is designed to share aggregated occurrence data with the public; access to non-aggregated data is limited to in-house use and can be made available on request to risk assessors. Hazard data from OpenFoodTox on the other hand can be accessed both via a web interface for quick screening of information and through a dedicated Zenodo repository for bulk download and direct reuse (e.g. in exposure calculations for risk assessments). In addition, data is version controlled and linked to persistent citable DOIs. This, in our view, strongly facilitates the timely dissemination of information and the reproducibility of the data analysis performed. The benefits of publishing data on an open repository such as Zenodo sparked a discussion at the IMR on how seafood data could be made available to a wider audience in the future, which is an important first step towards further implementation of the FAIR data principles.

2.3.3. Spin off activities in implementing FAIR data management practices

In addition to the work described above, during the project period supporting activities were carried out to improve communication in project work relying on coding and data sharing across different work groups and institutes. Within the Marine Toxicology group at the IMR several software tools are used to advance work on several cross-disciplinary projects. Microsoft Teams is used to communicate between members of the group through video calls or chat. Microsoft SharePoint is used as a document repository and for interactive document creation and editing. Linking SharePoint to OneDrive, within the group R code could be developed locally using RStudio. This allowed for efficient local collaboration between members of the team. To share different elements for a project externally, in addition to Teams, GitHub accounts were set up, and using RStudio scripts created earlier, were directly uploaded. This workflow was shown to VKM, which implemented this workflow, for the 'Risk-benefit assessment of sunscreen' (VKM et al., 2022b), that for the first time shares supplementary codes and datasheets interactively on Zenodo (Norwegian Scientific Committee for Food and Environment (VKM), 2022). Lastly, the fellow also engaged in discussions with IMR IT staff about modern software development and recent developments in micro-services architectures with standardised and structured data representation formats for sharing information between systems and services such as JSON and XML.
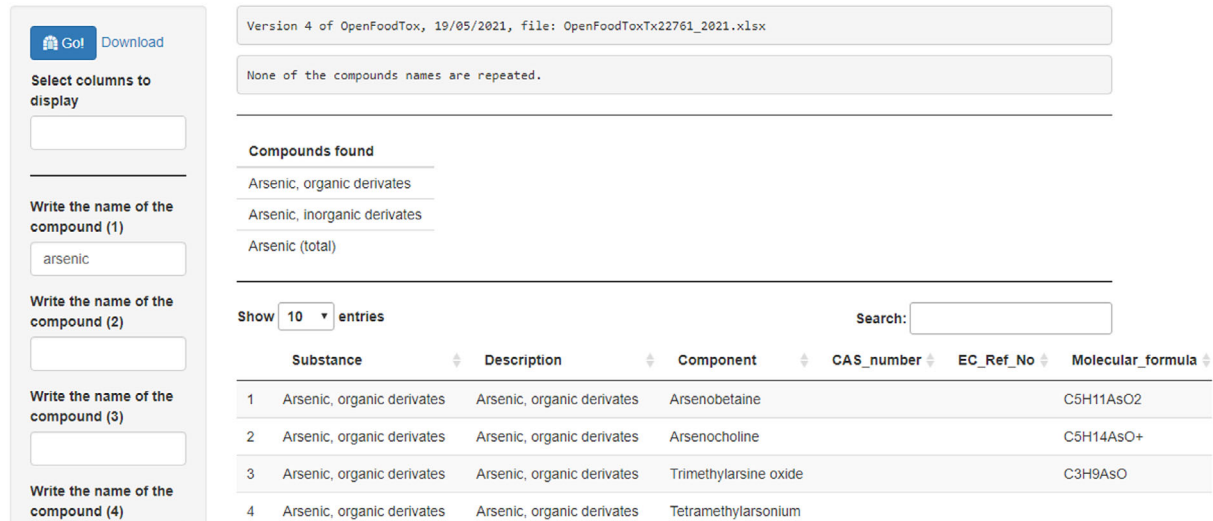
2.3.4. Other spin off activities during the project

The performed work is not only represented in this report but was also presented on a poster at the 'ONE Health, Environment, Society' conference in Brussels in June 2022. This conference also invites to participate in a video contest, where a short summary of the project was also presented. In addition to the project work related to FAIR data management, the EU-FORA programme also has offered further opportunities. Being integrated into the working group, the fellow had the opportunity to familiarise himself with a new field, participating in a paper regarding proteomics. Finally, to continue training in food security, the fellow also carried out the training 'Risk assessment in biotechnology' offered by the European Commission as part of the training initiative 'Better Training for Safer Food' (BTSF).

3. Conclusions

Large amounts of data that could be used in food safety risk assessments are available in different database. However, this steady increase of data has not been always followed by an improvement in the ways to easily access these data, data traceability or easy data reuse. To tackle these challenges, it is recommended that data for risk assessments must follow the FAIR principles, i.e. data must be Findable, Accessible, Interoperable and Reusable. Based on publicly available databases and open-source software tools, this project has been attempting to provide a proof of concept to show how using custom-made codes and alternative approaches could improve some characteristics of well-known databases including the OpenFoodTox Tool and the Seafood database. The use of platforms such as GitHub or Zenodo could make the data more Findable and Interoperable. The creation of web applications with Shiny could increase the accessibility to the data and make easy interaction between databases. The reusability was obtained through the selection of the appropriate formats for the data downloaded and the application of adequate systems to ensure traceability. Following these FAIR principles in the different databases is an essential step to ensuring the success of the future risk–benefit assessment timely and with adequate spending of human and economic resources.

Figure 1: User interface of the application designed with Shiny to access and work with the OpenFoodTox database

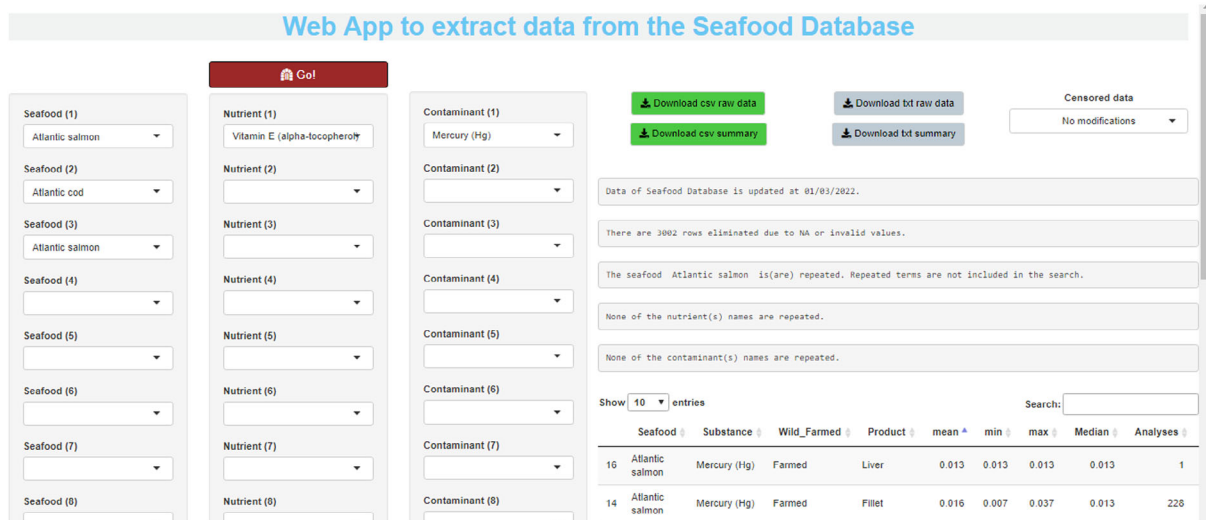**Figure 2:** User interface of the application designed with Shiny to access and work with the Seafood database

Figure 3: Results of the search in the Shiny application designed to work with the Seafood database. (A) Version of the database used, number of registers eliminated for errors and control of repeated inputs. (B) Examples of the summary and non-aggregated tables. (C) Options to download the results and the name of the file. (D) Options to control left-censored data

References

Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A and Borges B, 2021. Shiny: web application framework for R. R package version, 1, 1. Available online: https://CRAN.R-project.org/package=shiny

Dorne JLCM, Richardson J, Livaniou A, Carnesecchi E, Ceriani L, Baldin R, Kovarich S, Pavan M, Saouter E, Biganzoli F, Pasinato L, Zare Jeddi M, Robinson TP, Kass GEN, Liem AKD, Toropov AA, Toropova AP, Yang C, Tarkhov A, Georgiadis N, Di Nicola MR, Mostrag A, Verhagen H, Roncaglioni A, Benfenati E and Bassan A, 2020. EFSA's OpenFoodTox: an open source toxicological database on chemicals in food and feed and ists future developments. Environmental International, 146, 106293. https://doi.org/10.1016/j.envint.2020.106293

EFSA (European Food Safety Authority), 2019. Chemical Monitoring Reporting Guideline (SSD2). EFSA Journal 2019. https://doi.org/10.5281/zenodo.2543210

European Organization for Nuclear Research, & OpenAIRE, 2013. Zenodo. CERN, URS. https://doi.org/10.25495/7GXK-RD71

Giorgi FM, Ceraolo C and Mercatelli D, 2022. The R language: an engine for bioinformatics and data science. Life, 12, 648. https://doi.org/10.3390/life12050648

Hackenberger BK, 2020. R software: unfriendly but probably the best. Croatian Medical Journal, 61, 66–68. https://doi.org/10.3325/cmj.2020.61.66

Institute of Marine Research. 2022. Seafood database. Available online: https://sjomatdata.hi.no/ [Accessed: 1 March 2022].

Kovarich S, Ciacci A, Baldin R, Roncaglioni A, Mostrag A, Tarkhov A, Carnesecchi E, Gibin D, Di Piazza G, Pasinato L, Sartori L, Benfenati E, Yang C, Richardson J and Dorne JL, 2022. OpenFoodTox: EFSA's chemical hazards database. Zenodo. https://doi.org/10.5281/zenodo.780543

Microsoft. 2022. GitHub. Available online: https://github.com/ [Accessed: 21 July 2022].

Norwegian Scientific Committee for Food and Environment (VKM), 2022. Datasets and R-codes complementing the Opinion of the Norwegian Scientific Committee for Food and Environment (VKM) "Risk-Benefit Assessment of Sunscreen" [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6414372

Pineda-Pampliega J, 2022. EU-FORA Project. Available online: https://github.com/J-Pineda-Pampliega/EU_FORA_Project [Accessed: 28 July 2022].

R Core Team, 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.r-project.org/

RStudio Team, 2022. RStudio: integrated development environment for R. RStudio, PBC, Boston, MA. Available online: http://www.rstudio.com/

VKM, Andersen LF, Berstad P, Bukhvalova B, Carlsen M, Dahl L, Goksøyr A, Jakobsen LS, Knutsen HK, Kvestad I, Lillegaard IT, Mangschou B, Meyer H, Parr CL, Rakkestad KE, Rasinger J, Sengupta S, Skeie G, Starrfelt J, Ulven SM, Agdestein A, Bodin J, Bruzell E, Elvevoll E, Hessen DO, Husøy T, Krogdahl Å, Nilsen AM, Skjerdal T, Steffensen I-L, Strand TA, Velle G, Wasteson Y and Alexander J, 2022a. Benefit and risk assessment of fish in the Norwegian diet. Scientific Opinion of the Scientific Steering Committee of the Norwegian Scientific Committee for Food and Environment. VKM Report 2022. Norwegian Scientific Committee for Food and Environment (VKM). 17, Oslo, Norway. ISBN: 978-82-8259-392-2, ISSN: 2535-4019.

VKM, Bruzell E, Carlsen MH, Caspersen IH, Denison EM-L, Devold TG, Granum B, Mathisen GH, Pran ER, Rasinger JD, Rohloff J, Svendsen CH, 2022b. Risk-benefit assessment of sunscreen. Panel on Food Additives, Flavourings, Processing Aids, Materials in Contact with Food, and Cosmetics. VKM Report 2022, 10, ISBN: 978-82-8259-384-7, ISSN: 2535-4019. Norwegian Scientific Committee for Food and Environment (VKM), Oslo, Norway.

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J and Mons B, 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Abbreviations

DOI	digital object identifier
IMR	Institute of Marine Research
NAM	new approach methodology
VKM	Norwegian Scientific Committee for Food and Environment