

APPROVED: 29 May 2019

doi: 10.2903/j.efsa.2019.e170704

## Managing evidence in food safety and nutrition

Ermanno Cavalli<sup>1</sup>, Mary Gilsean<sup>1</sup>, Jane Van Doren<sup>2</sup>, Danica Grahek-Ogden<sup>3</sup>, Jane Richardson<sup>1</sup>, Fabrizio Abbinante<sup>1</sup>, Claudia Cascio<sup>1</sup>, Paul Devalier<sup>1</sup>, Nikolai Brun<sup>4</sup>, Igor Linkov<sup>5</sup>, Kathleen Marchal<sup>6</sup>, Bette Meek<sup>7</sup>, Claudia Pagliari<sup>8</sup>, Irene Pasquetto<sup>9</sup>, Peter Pirolli<sup>10</sup>, Steven Sloman<sup>11</sup>, Lazaros Tossounidis<sup>12</sup>, Elisabeth Waigmann<sup>1</sup>, Holger Schünemann<sup>13</sup> and Hans Verhagen<sup>1</sup>

<sup>1</sup>European Food Safety Authority (EFSA), IT; <sup>2</sup>Food and Drug Administration (FDA), USA; <sup>3</sup>Norwegian Scientific Committee for Food Safety (VKM), NO; <sup>4</sup>Medical Evaluation and Biostatistics Division, Danish Medicine Agency (DMA), DK; <sup>5</sup>Carnegie Mellon University, USA; <sup>6</sup>Department of Plant Biotechnology and Bioinformatics, University of Leuven, BE; <sup>7</sup>University of Ottawa, CDN; <sup>8</sup>Global eHealth Programme, University of Edinburgh, UK; <sup>9</sup>Harvard University, USA; <sup>10</sup>Florida Institute for Human and Machine Cognition, USA; <sup>11</sup>Cognitive, Linguistic, & Psychological Sciences, Brown University, CDN; <sup>12</sup>Inter-Agency Network for Information & Communication Technology (ICTAC), EC; <sup>13</sup>McMaster University, CDN

### Abstract

Evidence ('data') is at the heart of EFSA's 2020 Strategy and is addressed in three of its operational objectives: (1) adopt an open data approach, (2) improve data interoperability to facilitate data exchange, and (3) migrate towards structured scientific data. As the generation and availability of data have increased exponentially in the last decade, potentially providing a much larger evidence base for risk assessments, it is envisaged that the acquisition and management of evidence to support future food safety risk assessments will be a dominant feature of EFSA's future strategy. During the breakout session on 'Managing evidence' of EFSA's third Scientific Conference 'Science, Food, Society', current challenges and future developments were discussed in evidence management applied to food safety risk assessment, accounting for the increased volume of evidence available as well as the increased IT capabilities to access and analyse it. This paper reports on presentations given and discussions held during the session, which were centred around the following three main topics: (1) (big) data availability and (big) data connection, (2) problem formulation and (3) evidence integration.

© 2019 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

**Keywords:** data, evidence integration, evidence management, food safety

**Correspondence:** Ermanno.Cavalli@efsa.europa.eu

**Acknowledgements:** The European Food Safety Authority (EFSA) and authors wish to thank the participants of the breakout session '*Managing Evidence*' at EFSA's third Scientific Conference '*Science, Food and Society*' (Parma, Italy, 18–21 September 2018) for their active and valuable contribution to the discussion.

**Disclaimer:** The views or positions expressed in this article do not necessarily represent in legal terms the official position of the European Food Safety Authority (EFSA) nor of the US Environmental Protection Agency (US EPA). EFSA assumes no responsibility or liability for any errors or inaccuracies that may appear. This article does not disclose any confidential information or data. Mention of proprietary products is solely for the purpose of providing specific information and does not constitute an endorsement or a recommendation by EFSA for their use.

**Suggested citation:** Cavalli E, Gilsean M, Van Doren J, Grahek-Ogden D, Richardson J, Abbinante F, Cascio C, Devalier P, Brun N, Linkov I, Marchal K, Meek B, Pagliari C, Paschetto I, Pirolli P, Sloman S, Tossounidis L, Waigmann E, Schünemann H and Verhagen H, 2019. Managing evidence in food safety and nutrition EFSA Journal 2019;17(S1):e170704, 17 pp. <https://doi.org/10.2903/j.efsa.2019.e170704>

**ISSN:** 1831-4732

© 2019 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 License](https://creativecommons.org/licenses/by-nd/4.0/), 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
2. Thematic areas .....	5
2.1. Scientific innovation and new data streams .....	5
2.2. Distributed data: from 'data collection' to 'data connection' .....	5
2.3. Quantitative and data driven methods: transforming (big) data into scientific evidence.....	5
2.4. Exploring the living opinion: from static PDF documents to real-time analysis and communication .....	5
3. Summary of presentations .....	5
3.1. (Big) data availability and (big) data connection .....	6
3.1.1. The Fourth Revolution .....	6
3.1.2. The challenges of big data for European agencies .....	6
3.1.3. Blockchain technology and regulatory science.....	7
3.1.4. Software beats animal tests at predicting toxicity of chemicals.....	7
3.1.5. Managing data to manage evidence: social and technical challenges .....	8
3.2. Problem formulation .....	9
3.2.1. Ignorance and the community of knowledge .....	9
3.2.2. Human-computer sense-making models and the challenges of incorporating AI.....	9
3.2.3. Open source software paradigm: using ethics to promote technology and minimise its risks .....	9
3.3. Evidence integration .....	10
3.3.1. Integrating mechanistic evidence from evolving sources in hazard and risk assessment .....	10
3.3.2. Foodomics 2.0 .....	10
3.3.3. Network-based integration of molecular 'omics data.....	11
3.3.4. EFSA's Scientific Colloquium on 'omics in risk assessment .....	12
3.3.5. From 'WoE' to quantitative data integration.....	13
4. Conclusions.....	13
5. Recommendations .....	15
References.....	15
Abbreviations .....	17

## 1. Introduction

Evidence ('data') is at the heart of EFSA's 2020 Strategy (EFSA, 2015a). Within the framework of strategic objective 2, 'Widen EFSA's evidence base and maximise access to its data', EFSA is addressing three underpinning operational objectives: (1) adopt an open data approach, (2) improve data interoperability to facilitate data exchange and (3) migrate towards structured scientific data.

Much of the data and evidence considered by the European Food Safety Authority (EFSA) is being made publicly accessible via its scientific data warehouse (EFSA, online), knowledge junction (Zenodo, online), repositioned *EFSA Journal*<sup>1</sup> on Wiley as well as European open data portals (European Commission, online). EFSA's published outputs are now available as JATS XML, the international standard for journal articles. The Authority is also piloting migration from PDF dossier applications in the regulated product area towards electronic dossier submission and automatic publication of non-confidential information using structured formats based, as far as possible, on existing international standards to enable data access and reuse.

The Royal Society has recently recommended that 'scientists should communicate the data they collect and the models they create' (The Royal Society, 2012), so that scientific conclusions are intelligible and assessable, plus being reusable by other scientists: data used for scientific papers should be accessible, intelligible, assessable and usable.

The generation and availability of data have increased in the last decade, as well as the IT capabilities to access and analyse them, potentially providing a much large evidence base for future (food safety) risk assessments. It is widely acknowledged that some 90% of the data in general in the world today has been created in the last two years and about 75% of these data are unstructured (Marr, 2018).

To that end, a full-day session at EFSA's third Scientific Conference on 'Science, Food and Society' (Parma, Italy, 18–21 September 2018)<sup>2</sup> was dedicated to the topic of 'evidence management'. The session started from the increased availability of potential evidence, and then it progressed into understanding how to frame and formulate problems so that only relevant evidence is taken into consideration. The session closed with a discussion on how relevant complex evidence can be integrated into a meaningful result. The session developed from earlier EFSA discussions on (big) data and evidence (EFSA, 2019) and brought together multidisciplinary speakers from industry, academia, the European Commission and food safety authorities to discuss developments, opportunities and challenges relevant to the acquisition and management of evidence to support future food safety risk assessments. It is anticipated that the outcome of this session will help to inform the next EFSA Strategy 2021–2027 (on data and evidence management).

Here, we first briefly summarise the status of data management in EFSA (EFSA, 2019), and subsequently summarise presentations made and discussions held during the breakout session 'managing evidence'. These contributions were centred around the following three main topics: (1) (big) data availability and (big) data connection, (2) problem formulation and (3) evidence integration.

The first topic considered the current situation of the increasing availability of a wide range evidence, from a myriad of different sources, at very different levels of relevance and reliability, with a rapidly increasing level of automation in their generation. Key elements included having access to relevant data and its extraction, and relevance of continually changing evidence. More relevant technologies in area of scientific risk assessment were discussed: molecular biology, 'omics, big data, the internet of things. Blockchain technology was discussed as a promising technology for data traceability and traceability in the whole food chain.

The second topic aimed to underline the increasing importance of problem formulation in a context in which there are continually increasing amounts of data available, and a resulting need to focus on questions that are relevant. Problem formulation has a human and societal dimension, and is important to determine the acceptable residual level of uncertainty (e.g. Devos et al., 2019a,b).

The third topic dealt with the weight of evidence assessment including evidence integration and uncertainty analysis and explored the applications of machine learning and artificial intelligence (AI) in scientific risk assessments.

<sup>1</sup> <https://efsa.onlinelibrary.wiley.com/journal/18314732>

<sup>2</sup> All conference materials are available at <https://www.efsa.europa.eu/en/events/event/180918>

## 2. Thematic areas

### 2.1. Scientific innovation and new data streams

It is widely acknowledged that the vast majority of world-wide data has been created in recent years and is unstructured. With this trend increasing exponentially, the data available for analysis will expand continuously in terms of all the 'Vs' of big data, including velocity, volume and variety (Gartner, online), and veracity and value (see below). EFSA is actively engaged in working with the scientific community to identify how data are created and captured in innovative research and what new and currently unknown or unexplored data streams EFSA should be accessing to further expand its evidence base (EFSA, 2019).

### 2.2. Distributed data: from 'data collection' to 'data connection'

Increasingly, the nature of EFSA's scientific work requires access to data not traditionally collected by the agency and it is timely to consider a shift in focus from 'data collection' to 'data connection'. An Application Processing Interface (API) is the back-end technology to facilitate this transition – effectively an electronic 'shop front' to EFSA's data for machines. Exploration of mechanisms to automatically connect to and retrieve data from outside the EFSA is a logical next step. Ultimately, an ecosystem of APIs has the potential to provide EFSA with access to up-to-date, relevant data without duplication and storage over-heads. Each data creator in the ecosystem could collect, validate, store, maintain and operate appropriate access controls. In addition, the availability of cloud computing will enable more effective processing of data (EFSA, 2019).

### 2.3. Quantitative and data driven methods: transforming (big) data into scientific evidence

Having identified and accessed relevant new data sources, the next challenge will be to ensure that the data quality (fitness-for-purpose) is appropriate to meet EFSA's standards of scientific rigour (Devos et al., 2019a,b). Building on the conclusions of the EFSA Prometheus report (EFSA, 2015b), quantitative methods for data appraisal and validation will need to continue to develop in parallel with changing approaches to data identification and retrieval to ensure its appropriate transformation into sound scientific evidence. EFSA's work in advancing its approach to automation, machine learning and AI will build on ongoing work towards interoperability standards and domain ontologies. There are considerable opportunities to work in collaboration with stakeholders to test complex predictive models and machine learning in risk assessment, expert knowledge elicitation and the tracing of food contamination events throughout the food chain (EFSA, 2019).

### 2.4. Exploring the living opinion: from static PDF documents to real-time analysis and communication

A key feature of data use in risk assessment is timeliness and this will continue to increase in importance in the future. Moving towards more real-time data analysis and risk communication is expected to be increasingly important. Making that analysis readily accessible to practitioners, scientists and consumers in a reproducible and transparent way will also be driven by modern data visualisation and dissemination services (EFSA, 2019).

## 3. Summary of presentations

The breakout session was designed to provide information on current challenges and future developments in evidence management within EFSA and food safety risk assessment, and to explore the associated level of automation in the above-mentioned thematic areas. The session considered three main aspects:

- (big) data availability and (big) data connection;
- problem formulation;
- evidence integration.

These aspects are currently handled by humans and/or by machines. Problem formulation is expected to remain largely carried out by humans. Evidence integration is currently mostly based on

human input, but it is expected to be increasingly addressed electronically. (Big) data availability and (big) data connection are already mainly managed electronically.

The first topic has been analysed in its different dimensions, from both theoretical and practical point of views: the Fourth Revolution in mankind evolution that has taken us into the data age; a practical example how big data can be used for evidence-based decisions by regulatory decision making; a practical example on blockchain as a technology that leads us into improved trust in evidence management; a practical example of successful application of software to reduce animal testing at predicting toxicity of chemicals; the social and technical challenges in sharing and reusing big data.

The second topic has been analysed in both its human and societal dimensions: the importance of crowdsourcing and communities of knowledge in contributing to societal problems; how machines can augment human performance via sense-making models; how open source software can result as a valid alternative in times of shrinking resources.

The third topic has been explored through analysis of 'omics disciplines and technologies, and through approaches as mode of action (MOA) and adverse outcome pathways (AOP), and methodologies as weight of evidence (WoE) and multicriteria decision analysis (MCDA).

### 3.1. (Big) data availability and (big) data connection

#### 3.1.1. The Fourth Revolution

The session opened with a presentation on 'the Fourth Revolution': after the first three industrial revolutions (steam engine, electricity and Internet), the Fourth Revolution is the explosion of data. Science for risk assessment makes huge strides in the age of big data and high-performance computing. Science can feed ever clearer and faster risk assessment, and real-world – and real-time – evidence from dynamic data can both improve the iterative maintenance of risk knowledge and create opportunities for more efficient and effective management steps.

These revolutionary changes affect a world where the human mind evolves at a different rate. The evolution gap drives much of the cognitive dissonance, distrust and dystopia around the ongoing Fourth Revolution. In the upcoming data age, science is lost as a driver of democratic policy, when there are no new frameworks for society as it confronts risk. With modest adjustment, the existing model can guide us into the coming decades, with Europe still being a creative and respected thought leader in the global risk community, provided that:

- by re-engaging with what we know about human society's needs in terms of explanation and empowerment, as well as risk management, institutions can better fulfil their ever-legitimate roles;
- by focusing on the outcomes for which food safety systems stand, scientific evidence can also guide and underpin the changes in method and tools that offer better results with lower 'compliance friction'.

Successful data futures (for EFSA, and for Europe in general) should be based on Truth and Science, Broad Goals, Trustworthy Processes, Fullest Participation, Innovative Mindsets and the right tools (EPSC et al., 2016), along with the five Vs for data.

As such, the data age is characterised by (Olson, online):

- volume (public cloud will be needed, to store amounts of data that will exponentially grow);
- variety (requiring AI and neural learning, to understand and cross-reference data coming from different and ever changing sources);
- velocity (5G internet, which continuously senses data from the real world); and by uncertainty about data's;
- veracity (requiring cybersecurity, distributed ledger technology (blockchain) and audit-bots);
- value (by asking the user, all the time).

#### 3.1.2. The challenges of big data for European agencies

As an example, the EMA/HMA big data task force is working on 'The challenges of big data for European agencies' (Heads of Medicines Agencies and European Medicines Agency, 2019). Big data, e.g. data collected in real time from the real world, in large amounts are increasingly becoming available for regulatory decision makers.



Main tasks of the task force are: to map relevant sources of big data in its remit (completed December 2017); to describe the current state, future state and challenges for regulatory expertise, competencies, need to specify legislation and guidelines, data analysing tools and systems; and, eventually, to generate a list of recommendations.

The path from real-world *data* to real-world *evidence* is, however, full of challenges, but has also many opportunities to contribute to evidence-based decisions (e.g. for the benefit of patients). Agreement on data standards, quality, linkage to clinical outcomes, as well as interoperability, are key challenges to address as we move forward.

### 3.1.3. Blockchain technology and regulatory science

The global food industry faces important governance challenges for which blockchain has been proposed as a solution. The concept of distributed ledger technology has already been identified for its value proposition from the perspective of different players in the food supply chain (producer, distributor, retailer, consumer, regulator, enforcer, etc.). The importance of separating the truth from the hype surrounding blockchain has been emphasised and identifying real-world examples that show how it is being used to promote transparency, traceability, accountability, reduce fraud and improve consumer trust. However, a note of caution is needed by examining some of the ways in which blockchain as a new 'trust machine' could potentially be subverted by vested interests and considering what the industry and regulators might be able to carry out to prevent this, as well as how innovations such as AI and social machines might help.

When asking if blockchain can help regulatory science, several promising areas of research and innovative practice were highlighted, as well as new opportunities:

- *Food and feed supply chain*: Blockchain can be used to uniquely identify and track the provenance, stewardship and condition of Food and Feed at every step in the supply chain.
- *Detect food substitution*: Recent reports suggest that 'up to a fifth of meat tests in the UK reveal unspecified DNA' (Mackay, 2018) and that 41% of seafood samples tested in Canada were mislabelled (Ruryk and Chung, 2018). Using blockchain to create a verifiable ledger of food components and ingredients can help to prevent fraud by making it easier to identify changes. Integrating biometrical data also holds promise: for example, Australia is currently exploring how to record 'chemical barcodes' on blockchain to combat food counterfeiting.
- *Preventing food-borne disease*: Walmart is leading an effort, together with Nestlé, Dole and others, to seek a Farm-to-Grocery-Aisle View of the Food Supply Chain. Early results show that it went from needing 6 days to 2 s to trace a contaminated product from shelf to source, after implementing blockchain. It was explained how improving both the speed and precision of source identification also benefits farmers, by avoiding unnecessary food destruction and stigma, while also helping to build consumer trust and reduce costs.
- *Real-time regulatory compliance*: Companies such as Modum<sup>3</sup> are coupling the internet of things and cloud-based digital ledger approaches to monitor temperature-sensitive products during freight transportation, automatically auditing compliance with regulations on storage conditions, while helping to avoid transport-related spoilage.
- *Scientific currency*: Some of the future ways in which blockchain could be used to support science and innovation include tracking unpublished negative research to aid open science for better decision making, enabling academics to record their microcontributions to emerging research or impact, and auditing incremental innovations that challenge regulatory reclassification.

### 3.1.4. Software beats animal tests at predicting toxicity of chemicals

On AI, it was recently found that 'software beats animal testing at predicting toxicity of chemicals' using data from an EU agency (Hartung, 2019). Nowadays much chemical safety information is publicly available, raising the hope to facilitate the prediction of the toxicity of new chemicals. The REACH Regulation (1907/2006) provided the possibility of using non-testing opportunities, including Quantitative Structure–Activity Relationship (QSAR) and Read-across (RAx). RAx is the approach to estimate the toxicity of a substance by comparison to known properties of similar substances. By

---

<sup>3</sup> <https://modum.io/>

analysing the data submitted after the first REACH deadlines in 2010 and 2013, RAx resulted as the predominant method to avoid new tests on animals.

This triggered new general initiatives with the common goal of improving the robustness of the approach and the scientific endorsement for better regulatory acceptance. For example, the European Chemicals Agency (ECHA) has published their Read-Across Assessment Framework (RAAF) to guide the formal presentation of RAx data and the European project EU-ToxRisk<sup>4</sup> works on eight case studies on RAx with the engagement of industry and regulators.

The Center for Alternatives to Animal Testing (CAAT) in the USA and Europe has started a RAx programme in 2014 to develop Good Read-Across Practice and through the analyses of the data in the ECHA database of registered substances. At that time, the number of substances was about 10,000 and after transformation of the records into a computer-readable format, it was possible to improve toxicity predictions exploiting the big data available. Analysing the six most common OECD toxicity tests for reproducibility (consuming 55% of animals in safety testing in Europe), a reproducibility of 81% (balanced accuracy, sensitivity 70%) was found.

Today the system has expanded with data from PubMed and the US National Toxicology Programme to 10 million structures, including 800,000 chemicals with millions of data points on physicochemical and toxicological endpoints. The constructed models automate and extend the regulatory assessment (RA) method of chemical classification. The new approach, called read-across structure-activity relationship (RASAR) uses machine learning to combine binary fingerprints and Jaccard distance to define chemical similarity and feature vectors for supervised learning. A boost in predictivity was achieved by data fusion, i.e. each prediction was deduced from 74 different features.

The results are focused on nine endpoints: skin/eye irritation, skin sensitisation, acute oral/dermal/inhalation toxicity, mutagenicity and acute/chronic aquatic toxicity. Predicting 190,000 toxicity classifications of chemicals, this system provided high sensitivities and specificities that equal or even surpass the predictivity of animal tests. For the six tests, for which animal study reproducibility could be assessed, the RASAR resulted in 87% accuracy (sensitivity 89%).

Overall the work demonstrated that toxicological studies carried out using animal testing are only 81% reproducible, and only 69% reproducible for toxic chemicals. The RASAR predictions, fully based on software and AI, obtained 87% (balanced) accuracy.

It is envisaged that similar results should be achieved in the realm of food safety: software should strongly reduce animal testing in this area as well. The next challenge is how far regulatory bodies and governmental agencies will promote and encourage such new possibilities offered by big data.

### 3.1.5. Managing data to manage evidence: social and technical challenges

On the route from managing data to managing evidence, both social and technical challenges appear, such as the reuse of open research data. In this context, 'reuse' simply refers to the usage of a data set released in open access by someone other than the originator. While calls for increasing data sharing and interoperability have recently gained traction among science stakeholders, little information is known about how scientists actually reuse data once these have been made available in open access. Expectations for how much, how and by whom open data will be reused are often misplaced.

Research shows that, on average, among the many data sets a data repository can host, just a few popular open data sets are highly reused, while most data sets are very rarely reused, or not at all (this is also called 'the long tail of data reuse'; Wallis et al., 2013) and is comparable with the citation of scientific literature in general. Data curation (i.e. documenting and organising data into structured formats) is necessary but not sufficient for reuse: scientists reuse solely those data that are instrumental to their own research agenda and workflows. Experimental data (as opposed to observational data) are the hardest to reuse: the epistemic costs of learning about the data and the science behind it are often too high, specialised knowledge cannot be easily formalised in metadata and ontologies.

Science open data can be reused for many goals. Scientists commonly reuse these for control, comparison, or calibration, or (more rarely) to conduct meta-analyses and to train or test algorithms. However, setting aside a few notable exceptions, scientists almost never reuse open data sets to investigate novel research questions (i.e. for knowledge discovery). Scientists easily trust open data that have been reused before, over and over again, by their colleagues. Newly released open data with no record of reuse will need time to conquer scientists' hearts, the adoption curve could take months, or even years. Reputation, trust and pre-existing networks impact reuse as much curation practices.

---

<sup>4</sup> <http://www.eu-toxrisk.eu/>



Communities of practice are the ultimate holy grail of data reuse. For all the reasons mentioned above, it was pointed out that the most successful cases of data reuse originate from multilaboratory collaborations that involve both the data creators and the new users.

## 3.2. Problem formulation

### 3.2.1. Ignorance and the community of knowledge

Asking people to explain how something works reveals an illusion of explanatory depth: typically, people know less about how things work than they think they do. We over-estimate our knowledge of common objects. We similarly overestimate our understanding of political policies. It was argued that this illusion of understanding relates to living in a community of knowledge, guided by shared intentionality. Our communities understand how things work and we fail to distinguish what we know from the knowledge that resides in other people's heads. What was drawn out was some of the implications of these ideas for the importance of ferreting out expertise from parts of society in which it resides.

### 3.2.2. Human-computer sense-making models and the challenges of incorporating AI

On problem formulation, human-computer sense-making models and the incorporation of AI are real challenges. Multilevel computational cognitive models can provide opportunities for understanding and engineering new systems for sense-making and decision making in complex domains, such as intelligence analysis. A multilevel model of the human sense-making process is necessary to develop optimally performing human-AI interaction systems. Such a model accounts for the different levels of processes (psychological, rational and organisational) that occur at different time scales. Such multilevel models have a number of benefits. They allow researchers to predict:

- how difficult it will be for someone to find certain information;
- how much a person will learn from using a particular system;
- whether people will be biased in their information searches and sense-making;
- what kinds of credibility judgements people will make about information sources.

As such, models of human cognition in sense-making have informed the design and engineering of higher performance systems. For the foreseeable future, AI components in complex sense-making tasks will not be autonomous, rather they will work interdependently with human specialists and mixed human-AI systems 'teams' pose new challenges and require new designs. Those designs can be informed by new cognitive science research focused specifically on human-AI interaction in sense-making tasks. A major challenge, going forward, is setting up successful human-AI interdependence. The emerging standard model of cognitive processing and AI is about a fairly constrained, reasonably well defined set of tasks and domains, but, the world we live in is not well defined, it is very open ended. Current systems of AI are extremely difficult to understand. They produce valuable predictions and behaviour, but we often have difficulty explaining how or why these were produced. This influences trust, as people tend to trust those things that they understand or are familiar with. Explainable AI is one of the major new fields of research relevant to sense-making and decision-making systems.

### 3.2.3. Open source software paradigm: using ethics to promote technology and minimise its risks

Free and open source software (FOSS) has been at the heart of the ongoing information age for more than three decades, to an extent that today it is impossible to underrate its value, ignore its influence or challenge its momentum. As such, there exists an 'open source software paradigm' and ethics can be used to promote technology and minimise its risks. FOSS has been perceived, during such period. Not only as the driving force of innovation, but also as an inspirational moral architecture, warranting the equitable sharing of benefits from the digital transformation of society.

An evidence-based approach was presented to showcase the overall potential and the perspectives offered by FOSS today. The central tenet of the presentation revolved around FOSS maturity, and several aspects were demonstrated, such as FOSS's readiness:

- to support effectively today's complex information architectures;
- to operate reliably critical business applications;
- to help reduce the risks of proprietary solutions by avoiding costly vendor lock-in.

The capabilities offered by FOSS today allow the co-shaping of new technologies such as AI, cloud computing and big data. Emerging FOSS products, new licensing modes and business models were presented, along with FOSS costs, benefits and risks. These were all with a view to understand the current FOSS potential at a time of transition and its present relevance in a context of financial, social and environmental crisis that public organisations are facing.

In this context, the FOSS initiatives in global organisations with emphasis on the EU bodies were overviewed. Their experience has been crystallised in a host of best practices and strategic policy initiatives, e.g. the digital single market, the new General Data Protection Regulation (GDPR), the digital workplace, the future of work etc. FOSS capabilities to support reliably public organisations at optimal cost, flexibility and significant innovation potential were highlighted. Strategic considerations and key drivers in selecting, building and operating FOSS solutions have been set in perspective with technology governance and policy implications for society and for the economy at large.

### 3.3. Evidence integration

#### 3.3.1. Integrating mechanistic evidence from evolving sources in hazard and risk assessment

Within integrating mechanistic evidence from evolving sources in hazard and risk assessment, modern concepts such as MOA and AOP are conceptually similar constructs that organise mechanistic knowledge as a sequence of measurable key events at different levels of biological organisation (Lanzoni et al., 2019). AOPs address chemically agnostic key events between the initial interaction of a chemical with a molecular target (the molecular initiating event (MIE)) and adverse or disease outcomes. MOA analysis for hazard characterisation includes additional consideration of the chemical-specific aspects of disposition to the target (toxicokinetics and metabolism).

These pathway descriptions facilitate integrating and assessing mechanistic data in hazard and risk assessment from a broad range of sources including structure–activity analysis, *in vitro* assays, toxicity tests in animals and observational or clinical studies in humans. Linkage of MIEs and early key events measured in higher throughput systems to adverse effects characterised in traditional testing strategies is also anticipated to advance more tailored, efficient and predictive testing strategies.

The development and description of AOPs has been formalised in a public knowledge base to support their use for various applications in testing and assessment within an OECD programme. The associated Guidance and Users Handbook outlines conventions, terminology and relevant information content, including a structured assessment of the extent of supporting evidence. WoE is characterised based on a subset of the Bradford Hill (B/H) considerations applied to assess causality in epidemiological studies and tailored, more recently, for application to mechanistic data in international frameworks for MOA analysis. Examples of the nature of data sets associated with high, moderate and low confidence for the defined considerations were provided. The considerations have also been rank ordered to reflect their relative importance and the extension of the approach to enable quantitation of comparative WoE is being considered.

This structured and systematic consideration of the extent of mechanistic evidence in a coordinated construct, such as the AOP, facilitates its use for various regulatory applications for which different degrees of confidence are required. Coordinated consideration early in AOP development also focuses research efforts to meet specific regulatory need, based on critical data gaps.

#### 3.3.2. Foodomics 2.0

'Omics have found their way into food and nutrition too. 'Foodomics 2.0' was defined a decade ago as 'a discipline that studies the food and nutrition domains through the application and integration of advanced 'omics technologies to improve consumer's well-being, health and knowledge'. Foodomics 2.0 is warranted as all of the underlying technologies have had significant upgrades over the past years. Molecular characterisation is currently available at an unprecedented scale. In addition, downstream (big) data analysis had evolved and adopted insights/methods from fields ranging from information and communications technology (ICT) (computer science and database technology) through statistics/machine learning and AI. In many ways our biggest bottleneck is our ability to draft realistic, interesting, but non-trivial questions around nutrition and health. A plethora of current and (near) future applications of molecular profiling in food sciences has been presented as Foodomics 2.0 is getting ready for personalised nutrition.

Nutri(epi)genomics is 'a discipline that studies the food and nutrition domains through the application and integration of advanced 'omics technologies to improve consumer's well-being, health and knowledge'. Some examples were described on declination of foodomics in different domains, including: (i) genetics (nutrigenomics); (ii) epigenetics (nutri-epigenetics); (iii) metagenomics (nutri-metagenomics); and (iv) examples in applied technology on big data/AI.

As concerns genetics and nutrigenomics, 5 years ago it was already possible to have access to services for creation of a personal protein profile in a digital file, and even to have a 3D printing of one's own proteins. In 2017, the consumer genomics market expanded rapidly with genetic testing able to offer to consumers a number of tests spanning ancestry to love finding. When questioning what we actually can do with consumer genetics, Wobblebase's<sup>5</sup> work bridges the gap between consumers and medical professionals to leverage genetic information to its fullest potential. For instance, an app has been developed that is able to identify pills with a phone camera, unravel their ingredients and provide the user with a report on their own single nucleotide polymorphisms (SNPs) and sensitivity to medicines, based on interrogation of pharmacogenomics databases. In nutrition, a similar approach can be used to scan barcodes on food, identify the ingredients and check consumer's SNPs against sensitivity to ingredients such as peanuts or to allergens.

Epigenetics makes it possible to reuse one genome for many different purposes and is driving the aetiology of many human diseases, and we can influence changes in the epigenetics. The relative importance of epigenetics is that food is inducing changes in DNA (expression) continuously! There is an integration of intrinsic and environmental signals: food can switch-on or switch-off genetics and so modulate its functioning as it happens, e.g. queen bee development following nutrition with royal jelly. While the genetic code in DNA is not the target, changes in histones and methylation of DNA bases, for example, can modulate gene expression.

Metagenomics describe that a healthy adult hosts have ca. 100 trillion bacteria in the gut alone and the communal gut microbial genome (microbiome) is ca. 150 times larger than the human genome. It is reasonable to view the microbiome as some sort of an organ by itself, for which there are kits available on the market (uBiome<sup>6</sup>) that are able to measure metagenomics profiles using machine learning, AI, advanced statistical techniques and sequencing technologies.

There are already home version portable and very cheap laboratory equipment available on the market (Foodomics 3.0), such as a 'Minima' kit for consumers who want to test the origin of their food in 1 or 2 h at home. This kit is based on a DNA amplification technique via loop-mediated isothermal amplification (LAMP<sup>7</sup>). Such miniaturised portable, real-time sequencing devices are as small as a mobile phone and can be readily plugged into a laptop. Indeed, testing is getting closer to the user, and can be carried out at home.

Similarly, a large amount of information (big data) is available on the combination of foods. AI and text mining can be used to screen this information to survey food recommenders (Anderson, 2018) or to understand and exploit the science behind food pairing to discover new food pairs.<sup>8</sup>

### 3.3.3. Network-based integration of molecular 'omics data

Networks can facilitate the integration of molecular 'omics data. Interaction networks in which nodes represent biological entities, such as genes, gene products, metabolites, etc., and edges represent the interactions between the nodes, provide a comprehensive way of summarising all available molecular information known about an organism of interest. These networks can be used not only to visualise biological data, but can also serve as a scaffold to analyse one's own in-house generated data sets. Network models, built on prior interaction networks, provide an intuitive way to integrate heterogeneous 'omics data and demonstrate why network models are so powerful. An example was provided on how cohort analysis can be used to assess the efficiency of drug responses or to unveil the molecular mechanisms that result in antibiotic resistance development.

'Omics data can be used for food safety, for instance for assessing the molecular signature of a toxic compound. Transcriptomics can for instance show which genes are upregulated or downregulated in relation to exposure to a certain food. Expression profiles of individuals can be used as a molecular phenotype to identify molecular signatures that can be used as biomarkers and ideally unveil the MOA of toxicity. In personalised nutrigenomics, via genomic biomarkers, one could determine whether or

<sup>5</sup> <https://doc.ai/>

<sup>6</sup> <https://ubiome.com/>

<sup>7</sup> <http://minimabiolab.org/project/>

<sup>8</sup> <https://www.foodpairing.com/en/home>

not an individual will elicit a toxic response in response to a certain toxicant. Identifying genomic biomarkers requires 'omics-based cohort analysis.

The use of network-based analysis increases the possibility of detecting changes, as well as increasing the power of analysis. These networks exploit prior information on molecular interactions as a scaffold to drive the analysis, steering the solution of the data-integration problem to the most biologically relevant network. Networks basically provide an intuitive scaffold to integrate data. Network-based analysis of 'omics data can allow the search for pathways that contain significantly more variants in the responder set than in the non-responder set. Both single network models and complex network models can be used, using different data sources. Network-based analysis of 'omics data also can provide insight into mechanism of disease.

### 3.3.4. EFSA's Scientific Colloquium on 'omics in risk assessment

In research, 'omics data has been used for more than a decade to study basic biological problems and vast amounts of analytical data are being collected and shared in public database; however 'omics data and its approach are not (yet) used very much in risk assessment. 'Omics data sets are starting to be used in some risk assessment areas and are used as complementary to or as a substitute for classical data.

It is clear that 'omics have now entered the domain of food safety, nutrition and risk assessment. Already in 2014, EFSA started mapping the use of 'omics tools in risk assessment related to food and feed safety and to review modern methodologies and tools for human hazard assessment of chemicals. This process has continued with EFSA's 24th Scientific Colloquium on 'Omics in risk assessment: state of the art and next steps' that took place in Berlin, Germany on 24–25 April 2018 (for programme and presentations, see <http://www.efsa.europa.eu/en/events/event/180424-0>). The Scientific Colloquium aimed to take into account the latest advancements and explore the potential use of 'omics data sets to support scientific safety evaluation. It addressed the question 'are there concrete possibilities of implementation of 'omics in risk assessment?'

The outcome of the Colloquium intends to support risk assessors in the process of incorporating 'omics tools into the risk assessment of food and feed products. An event report summarising the outcomes of the Colloquium was published in 2018 (EFSA, 2018).

Outcomes of this event indicate that 'omics studies are used to characterise and quantify the roles and relationships of large sets of different types of molecules in an organism to collate information on the functional status or impact of environmental factors on an organism. In recent years, the development of innovative tools in genomics, transcriptomics, proteomics and metabolomics (designated collectively as 'omics technologies) has opened up new possibilities for applications in scientific research and led to the availability of vast amounts of analytical data. Omics are used to characterise and quantify the roles and relationships of large sets of different types of molecules in an organism. Genomics can facilitate the analysis of entire or component genome sequences of an organism. Transcriptomics and proteomics provide significant bodies of information on temporal and spatial expression of genes and gene products, respectively, while metabolomics captures data for a large pool of metabolites. The interpretation and integration of 'omics data can provide valuable information on the functional status of an organism and on the impact of external factors, e.g. stressors.

The Colloquium explored the opportunities for integration of data sets produced via specific 'omics tools within the remit of EFSA's risk assessment approaches and has built further towards a concrete path of implementation. Discussions in the Colloquium focused on a set of topics for which EFSA intends to exploit 'omics data sets to support scientific safety evaluation. These topics are: genomics in microbial strain characterisation; metabolomics for the comparative assessment of genetically modified (GM) plants; and the use of 'omics for toxicological and environmental risk assessment.

In summary, outcomes and discussion of the 24th Colloquium highlighted that:

- 'omics data are an important tool, e.g. in elucidating mechanisms and determining MOAs and AOPs;
- 'omics data can be integrated into risk assessment in several areas, albeit that there are development needs (reference data sets, information on baseline variability, quality and reporting standards, development of expertise);
- 'Omics data are already part of risk assessment data in EFSA (WGS data in analysis of food-borne diseases and for GM plants);

- challenges are linked to:
  - storage/cloud-based storage;
  - data analysis (development of software tools and expertise);
  - setting quality standards and developing guidelines;
  - interpretation in risk assessments and developing guidelines.

### 3.3.5. From 'WoE' to quantitative data integration

Finally, within evidence integration, the audience was informed on the topic 'From 'WoE' to quantitative data integration'. In line with the recent EFSA activity on WoE (EFSA Scientific Committee, 2017), WoE is an approach that, by means of qualitative or quantitative methods, integrates individual lines of evidence to form a conclusion. Contemporary WoE methodologies have advanced with the evolution of statistical science (see Linkov et al., 2015 for review). In the 1960s, it was proposed that WoE processes should follow an inherently Bayesian statistical approach in which 'prior' beliefs for or against a particular hypothesis are updated after evaluation of information or evidence to achieve a 'posterior' belief. It has evolved in multiple methods and tools with varying degree of quantitative rigour and reliance on judgement. From commonly used listing evidence and best professional judgement all the way through quantitative MCDA and Bayesian methods, all the methods are captured to a varying degree by recent WoE recommendations and approaches that have been expanding recently. Multiple applications of these methods for food safety and related areas have been reported and have been summarised in this presentation. It was argued that successful application of WoE to food safety requires standardisation to produce consistency and comparability across ongoing WoE efforts.

It is essential that the WoE methodology used in risk assessment moves from qualitative approaches that are commonly not used to quantitative analysis and decision making. Increased rigour of analysis can help to:

- determine if there is enough evidence to support a determination or action (e.g. justification for a specific threshold);
- compare alternative courses of action or selected alternative agents of processes to see what is better supported (e.g. selection of the most likely MOA);
- identify gaps in understanding and prioritise research;
- highlight scientific consensus to bolster use of an approach/tool.

Any selected WoE methodology needs to reflect the reason for the analysis, and MCDA methods (Cegan et al., 2017) can help here. The following aspects on MCDA methods are to be noted:

- MCDA methods have evolved as a response to the observed inability of people to effectively analyse multiple streams of dissimilar information;
- Many different MCDA approaches are based on different theoretical foundations (or combinations);
- MCDA methods provide a means of integrating various inputs with stakeholder/technical expert values and means of communicating model/monitoring outputs for regulation, planning and stakeholder understanding;
- MCDA approaches can guide collection of additional information through the value of information analyses;
- Risk-based MCDA offers an approach for organising and integrating varied types of information to perform rankings and to better inform decisions.

Several recent applications of quantitative WoE clearly show its value for regulatory decision making in the context of EFSA interests (Linkov et al., 2011; Becker et al., 2015; Collier et al., 2016).

## 4. Conclusions

In the final forum discussion, the findings and presentations of the day were summarised and analysed in depth.

Exploration of all plausible data streams, including from the general public, could generate useful information to inform future scientific work. To be able to source knowledge and information available in the public domain, existing IT platforms need to be explored or developed to facilitate the harvesting and exchange of data and ideas using crowdsourcing. An assessment of data quality (fitness for purpose) before use is a prerequisite in this regard.



While keeping in mind that the world is going through an exponential explosion of data, potentially increasing the evidence base for risk assessments, it becomes relevant to consider how we can explore and analyse all these data.

Besides the requirements of new tools and approaches such as machine learning and AI, the session addressed the challenges ahead with respect to targeting, analysing and accessing the right data serving as evidence, and reflecting on the ways in which this diverse body of evidence could be used to provide fit-for-purpose risk assessments.

We are now facing many sources of evidence, many sources of knowledge and many different ways of incorporating them. The trade-off between qualitative and quantitative risk assessment was discussed, and how big data can influence both. A residual level of uncertainty will always exist, both due in biases in defining a model, and possible non-uniform coverage in the collection of big data. It is as well forecasted that risk assessment will not be totally automated, and there will still be a residual role for expert judgement. An agreed point is that we need, in the future, a better way to record decision making, to improve quality and transparency.

It has been discussed how to avoid the apparent dichotomy between randomised trials and observations from the 'real world', and how the two can be integrated to decide on causality and improve evidence-based decisions.

The advantages and disadvantages of 'real-world' observations have been analysed, being conscious that different perspectives of the same reality can be observed, leading to different purposes of collection of big data, and therefore making data reuse difficult.

B/H criteria were discussed and considered as a possible decision model, while biological plausibility was discussed from several point of views, including being a criterion of coherence between theory and the 'real world'.

The panel did agree on the difficulty of integrating 'omics and AOP.

As a summary, the three main aspects of the session have been:

- (big) data availability and (big) data connection;
- importance of problem formulation;
- evidence integration.

The first topic considered the current situation of the increasing availability of a wide range of evidence.

Future areas of research should include: (i) volume: the need for an EFSA public cloud; (ii) variety: the need for AI and neural learning; (iii) velocity: the need for a 5G internet of food, that continuously senses data from the real world; (iv) veracity: the need for cybersecurity, distributed ledger technology (blockchain) and audit-bots; and (v) the need to determine and document value: we need to ask users.

More relevant technologies in the area of scientific risk assessment were explored: molecular biology; 'omics, big data, the internet of things.

Blockchain technology was explored as promising technology for data traceability and traceability in the whole food chain.

In human health risk assessment, the potential for software to increase the capacity to predict toxicity of some chemicals has been demonstrated.

The second topic underlines the increasing importance of problem formulation in a human and societal dimension.

The importance of crowdsourcing to aggregate diverse expertise has been demonstrated, along with the possible usage of blockchain as a sophisticated method for effectively sharing reputation and creating trust.

Sense-making has been discussed as aid in designing computer systems that effectively and sensibly augment the performance of humans.

The third topic deals with WoE assessment, including evidence integration and uncertainty analysis and explores the applications of machine learning and AI in scientific risk assessment.

'Omics combined with big data sets, large computational power and machine learning can be integrated into the risk assessment to cover several of areas, although there are development needs (e.g. reference data sets, information on baseline variability, quality and reporting standards, develop expertise). Moreover, several challenges need to be overcome in terms of storage (→ cloud-based storage), data analysis (→ software tools and expertise development), setting quality standards (→ guideline development) and interpretation for risk assessment (→ guideline development).

MOA and AOP are complementary disciplines to assimilate mechanistic data to move us from the observation in animal studies to more predictive approaches, by assimilating biochemical and biological



mechanistic information on disease pathways at a broad range of biological levels of organisation. This can be extremely useful for a range of regulatory applications (e.g. development of testing strategies, considerations on biological plausibility in epidemiological studies, MOA analysis for specific chemicals or groups, environmental monitoring).

WoE methodology is essential to move from qualitative to quantitative analysis and decision making. It helps to: determine if there is enough evidence to support a determination or action (threshold); compare alternatives to see what is better supported (carcinogen MOA); identify gaps in understanding; and highlight scientific consensus to bolster use of an approach/tool.

MCDAs methods have evolved as a response to the observed inability of people to effectively analyse multiple streams of dissimilar information. Many different MCDA approaches exist based on different theoretical foundations (or combinations). They provide a means of integrating various inputs with stakeholder/technical expert values, and of communicating model/monitoring outputs for regulation, planning and stakeholder understanding. Risk-based MCDA offers an approach for organising and integrating varied types of information to perform rankings and to better inform decisions.

## 5. Recommendations

The session has led to several recommendations from which EFSA can profit and which it can enrol in its future strategy (on data and evidence):

- To base EFSA's data futures on truth and science, broad goals, trustworthy processes, fullest participation, innovative mindsets and the right tools;
- The data age is characterised by the volume, variety and velocity of data generation, as well as their veracity and value;
- To further explore the strengths and limitations of the combined use of 'big data' and AI for risk assessment purposes, and consider development needs for its practical implementation;
- To further explore the strengths and limitations of biotechnological tools such as 'omics for risk assessment purposes, and consider development needs for their practical implementation;
- To further develop global partnerships to share and leverage data and analytical methodologies capturing, analysing and using these new types of data and data sources in risk assessment;
- To further explore the strengths and limitations of MOA and AOP approaches for risk assessment purposes, and consider development needs for their practical implementation;
- To further explore the strengths and limitations of the WoE methodology for risk assessment purposes, and consider development needs for its practical implementation;
- To further explore the strengths and limitations of MCDA methods for risk assessment purposes, and consider development needs for their practical implementation;
- To further explore the strengths and limitations of Blockchain technology for risk assessment purposes, and consider development needs for its practical implementation. Several promising areas of research have been identified:
  - scientific currency – blockchain technology as 'trust machine';
  - food/feed supply chain: blockchain technology can be used to uniquely identify every element of the food/feed supply chain;
  - food substitution: blockchain technology can be used to detect 'fakes' in the food chain;
  - preventing food-borne diseases: industries and wholesales are leading an effort to seek a farm-to-grocery-aisle view of the food supply chain.

## References

- Anderson, 2018. A survey of food recommenders.
- Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, Van der Burg B, Villeneuve DL, Watanabe H and Barton-Maclaren TS, 2015. Increasing scientific confidence in adverse outcome pathways: application of tailored Bradford Hill considerations for evaluating weight of evidence. *Regulatory Toxicology and Pharmacology*, 72, 514–537. <https://doi.org/10.1016/j.yrtph.2015.04.004>
- Cegan JC, Filion AM, Keisler JM and Linkov I, 2017. Trends and applications of multi-criteria decision analysis in environmental sciences: literature review. *Environment Systems & Decisions*, 37, 123–133. <https://doi.org/10.1007/s10669-017-9642-9>
- Collier ZA, Gust KA, Gonzalez-Morales B, Gong P, Wilbanks MS, Linkov I and Perkins EJ, 2016. A weight of evidence assessment approach for adverse outcome pathways. *Regulatory Toxicology and Pharmacology*, 75, 46–57. <https://doi.org/10.1016/j.yrtph.2015.12.014>

- Devos Y, Craig W, Devlin RH, Ippolito A, Leggatt RA, Romeis J, Shaw R, Svendsen C and Topping CJ, 2019a. Using problem formulation for fit-for-purpose pre-market environmental risk assessments of regulated stressors. *EFSA Journal* 2019;17(7):e170707, 17 pp. <https://doi.org/10.2903/j.efsa.2019.e170707>
- Devos Y, Elliott KC, Macdonald P, McComas K, Parrino L, Vrbos D, Robinson T, Spiegelhalter D and Gallani B, 2019b. Conducting fit-for-purpose food safety risk assessments. *EFSA Journal* 2019;17(7):e170708, 36 pp. <https://doi.org/10.2903/j.efsa.2019.e170708>
- EFSA (European Food Safety Authority), 2015a. EFSA Strategy 2020. Trusted science for safe food: Protecting consumers' health with independent scientific advice on the food chain. Parma, EFSA Available online: [https://www.efsa.europa.eu/sites/default/files/corporate\\_publications/files/strategy2020.pdf](https://www.efsa.europa.eu/sites/default/files/corporate_publications/files/strategy2020.pdf)
- EFSA (European Food Safety Authority), 2015b. Principles and process for dealing with data and evidence in scientific assessments. *EFSA Journal* 2015;13(5):4121, 35 pp. <https://doi.org/10.2903/j.efsa.2015.4121>
- EFSA (European Food Safety Authority), online. Data collection, standardisation and analysis. Available online: <https://www.efsa.europa.eu/en/science/data>
- EFSA (European Food Safety Authority), Aguilera J, Aguilera-Gomez M, Barrucci F, Cocconcelli PS, Davies H, Denslow N, Lou Dorne J, Grohmann L, Herman L, Hogstrand C, Kass GEN, Kille P, Kleter G, Nogué F, Plant NJ, Ramon M, Schoonjans R, Waigmann E and Wright MC, 2018. EFSA Scientific Colloquium 24 – 'Omics in risk assessment: state of the art and next steps. *EFSA Supporting Publications* 2018;15(11):EN-1512, 30 pp. <https://doi.org/10.2903/sp.efsa.2018.en-1512>
- EFSA (European Food Safety Authority), Cappè S, Gilsean M, O'Dea E, Richardson J and Verloo D, 2019. Editorial: The future of data in EFSA. *EFSA Journal* 2019;17(1):e17011, 3 pp. <https://doi.org/10.2903/j.efsa.2019.e17011>
- EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Benfenati E, Chaudhry QM, Craig P, Frampton G, Greiner M, Hart A, Hogstrand C, Lambre C, Luttk R, Makowski D, Siani A, Wahlstroem H, Aguilera J, Dorne J-L, Fernandez Dumont A, Hempen M, Valtueña Martínez S, Martino L, Smeraldi C, Terron A, Georgiadis N and Younes M, 2017. Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA Journal* 2017;15(8):4971, 69 pp. <https://doi.org/10.2903/j.efsa.2017.4971>
- EPSC, Madelin R and Ringrose D, 2016. Opportunity now: Europe's mission to innovate. European Union, Brussels. Available online: <https://publications.europa.eu/en/publication-detail/-/publication/a5d642ba-3f4e-11e6-af30-01aa75ed71a1/language-en>
- European Commission, online. EU Open Data Portal. Available online: [https://data.europa.eu/euodp/en/data/data-set?q=EFSA&ext\\_boolean=all](https://data.europa.eu/euodp/en/data/data-set?q=EFSA&ext_boolean=all)
- Gartner, online. IT glossary – big data. Available online: <https://www.gartner.com/it-glossary/big-data/>
- Hartung T, 2019. Predicting toxicity of chemicals: software beats animal testing. *EFSA Journal* 2019;17(7):e170710, 8 pp. <https://doi.org/10.2903/j.efsa.2019.e170710>
- Heads of Medicines Agencies and European Medicines Agency, 2019. HMA-EMA Joint Big Data Taskforce: Summary report. EMA, London. Available online: [http://www.hma.eu/fileadmin/dateien/HMA\\_joint/00-About\\_HMA/03-Working\\_Groups/Big\\_Data/2019\\_02\\_HMAEMA\\_Joint\\_Big\\_DataTaskforce\\_summary\\_report.pdf](http://www.hma.eu/fileadmin/dateien/HMA_joint/00-About_HMA/03-Working_Groups/Big_Data/2019_02_HMAEMA_Joint_Big_DataTaskforce_summary_report.pdf)
- Lanzoni A, Castoldi AF, Kass GEN, Terron A, De Seze G, Bal-Price A, Bois FY, Delclos KB, Doerge DR, Fritsche E, Halldorsson T, Kolossa-Gehring M, Hougaard Bennekou S, Koning F, Lampen A, Leist M, Mantus E, Rousselle C, Siegrist M, Steinberg P, Tritscher A, Van de Water B, Vineis P, Walker N, Wallace H, Whelan M and Younes M, 2019. Advancing human health risk assessment. *EFSA Journal* 2019;17(7):e170712, 23 pp. <https://doi.org/10.2903/j.efsa.2019.e170712>
- Linkov I, Welle P, Loney D, Tkachuk A, Canis L, Kim JB and Bridges T, 2011. Use of multicriteria decision analysis to support weight of evidence evaluation. *Risk Analysis*, 31, 1211–1225. <https://doi.org/10.1111/j.1539-6924.2011.01585.x>
- Linkov I, Massey O, Keisler J, Rusyn I and Hartung T, 2015. From 'weight of evidence' to quantitative data integration using multicriteria decision analysis and Bayesian Methods. *Altex-Alternatives to Animal Experimentation*, 32, 3–8. <https://doi.org/10.14573/altex.1412231>
- Mackay H, 2018. Meat testing: a fifth of samples reveal unspecified animals' DNA. *BBC News*, 5 September. Available online: <https://www.bbc.com/news/uk-45371852>
- Marr B, 2018. How much data do we create every day? The mind-blowing stats everyone should read. *Forbes*, 21 May. Available online: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6248809960ba>
- Olson M, online. What is big data?: Highlight from Opportunities Abound in the Big Data Space [video]. Available online: <https://ecorner.stanford.edu/in-brief/what-is-big-data/>
- Ruryk J and Chung E, 2018. Widespread mislabelling of seafood reported in cities across Canada. *CBC News*, 28 Aug. Available online: <https://www.cbc.ca/news/technology/seafood-mislabelling-fraud-1.4796762>
- The Royal Society, 2012. Science as an open enterprise. The Royal Society, London. Available online: <https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf>
- Wallis JC, Rolando E and Borgman CL, 2013. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE*, 8, e67332. <https://doi.org/10.1371/journal.pone.0067332>
- Zenodo, online. Knowledge Junction. Available online: <https://zenodo.org/communities/efsa-kj/?page=1&size=20>

## Abbreviations

AI	artificial intelligence
AOP	adverse outcome pathways
API	Application Processing Interface
B/H	Bradford Hill
CAAT	Center for Alternatives to Animal Testing
ECHA	European Chemicals Agency
EMA	European Medicines Agency
FOSS	Free and open source software
GDPR	General Data Protection Regulation
GM	genetically modified
HMA	Heads of Medicines Agencies
ICT	information and communications technology
LAMP	loop-mediated isothermal amplification
MCDA	multicriteria decision analysis
MIE	molecular initiating event
MOA	mode of action
QSAR	Quantitative Structure–Activity Relationship
RAAF	Read-Across Assessment Framework
RA	regulatory assessment
RASAR	read-across structure–activity relationship
RAX	Read-across
SNP	single nucleotide polymorphism
WoE	weight of evidence