SPECIAL ISSUE

EFSA Journal

doi: 10.2903/j.efsa.2016.s0509

Open risk assessment: data

Mary B. Gilsenan[1], Fabrizio Abbinante[1], Eileen O'Dea[2], Ana Canals[3] and Angelika Tritscher[4]

Abstract

Since its foundation, EFSA and the Member States have made significant progress in the area of data collection for risk assessment and monitoring. In partnership with competent authorities and research organisations in the Member States, EFSA has become a central hub of the European data on food consumption, chemical occurrence and foodborne outbreaks. Beyond EFSA's use of these data and sharing of contaminants and food consumption data with the World Health Organization and the Food and Agriculture Organization to support international risk assessment, they remain largely unexploited. In addition, for some of its risk assessments, EFSA also relies on published information, as well as on scientific studies sponsored and submitted by industry. The environment in which the Authority operates has significantly evolved since its foundation. The growth of digital technology has granted scientists and consumers alike faster and more efficient access to data and information. The open data movement, which has entered the sphere of the European Union institutions, is unleashing the potential for reuse of data. In parallel, the work of EFSA is increasingly subject to demands for more openness and transparency across its spectrum of stakeholders. EFSA aims to enhance the quality and transparency of its outputs by giving access to data and promoting the development of collaborative platforms in Europe and internationally. EFSA also plans to work with data providers and organisations funding research to adopt open data concepts and standards; gaining better access to, and making better use of, data from a wider evidence base. During the breakout session on 'Open Risk Assessment: Data' at the EFSA 2nd Scientific Conference 'Shaping the Future of Food Safety, Together' (Milan, Italy, 14–16 October 2015) opportunities and challenges associated with open data, data interoperability and data quality were discussed by sharing experiences from various sectors within and outside EFSA's remit. This paper provides an overview of the presentations and discussions during the breakout session.

Keywords: open data, interoperability, linked data, open science

Correspondence: editor-in-chief.efsajournal@efsa.europa.eu

Author affiliations: [1] European Food Safety Authority (EFSA), Italy; [2] Food Safety Authority of Ireland, Ireland; [3] Spanish Agency for Consumer Affairs, Food Safety and Nutrition, Spain; [4] World Health Organization, Switzerland.

Acknowledgements: EFSA wishes to acknowledge the contribution of the speakers at the breakout session on 'Open risk assessment: data' at the EFSA 2nd Scientific Conference 'Shaping the Future of Food Safety, Together' (Milan, Italy, 14–16 October 2015): Katy Börner, Indiana University; Jean-Claude Burgelman, European Commission; Leif Busk, Swedish National Food Agency; Stefano Cappè, European Food Safety Authority; Hans-Georg Eichler, European Medicines Agency; Britta Gallus, Metro Group; Yvo Volman, European Commission Publications Office; and Dave Weller, Thomson Reuters.

Suggested citation: Gilsenan MB, Abbinante F, O'Dea E, Canals A and Tritscher A, 2016. Special issue: open risk assessment: data. EFSA Journal 2016;14(S1):s0509, 16 pp. doi:10.2903/j.efsa.2016.s0509

ISSN: 1831-4732

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

Table of contents

1. Introduction

The European Food Safety Authority (EFSA) has become a central hub of the European data on food consumption and the occurrence of chemical and microbiological foodborne hazards. Some 14 million pesticide residue analytical records are transmitted to EFSA annually from monitoring programmes in European countries, as well as some 900,000 analytical records on contaminants in food and feed. EFSA also collates the European data on zoonoses, antimicrobial resistance (AMR) and foodborne outbreaks annually, as well as data on food additives. Some of these data are used to estimate dietary exposure for EFSA's risk assessments or for the compilation of annual European Union (EU) summary reports (i.e. on pesticide residues, zoonoses, AMR). Beyond EFSA's use of these data and sharing of contaminants and food consumption data with the World Health Organization and the Food and Agriculture Organization to support international risk assessment, they remain largely unexploited. In addition, for some risk assessments, in particular, those linked to premarket authorisation of regulated products, EFSA relies on published scientific information, as well as scientific studies sponsored and submitted by industry.

The open data movement is unleashing the potential for the generation of new knowledge and insights from reuse of data. In addition, the work of EFSA is increasingly subject to demands for more openness and transparency across its spectrum of stakeholders. Thus, it is timely to focus more attention on accessibility to the data in EFSA's data hub.

EFSA aims to be an advocate for openness by engaging with data providers to adopt open data concepts and standards within its risk assessment remit; in doing so, better access to and use of data from a wider evidence base will make risk assessment more robust and keep it relevant to scientific and technological progress, as well as societal concerns.

Here, we describe the proceedings of a breakout session on 'Open Risk Assessment: Data' at the EFSA 2nd Scientific Conference 'Shaping the Future of Food Safety, Together'. The objective was to bring together multidisciplinary speakers from industry, academia, the European Commission, the EU agencies and the Member States to share, discuss and debate opportunities and challenges associated with open data, data interoperability and data quality.

1.1. European Commission's Open Science Initiative: co-creating added value with data

An open approach to research data is core to the European Commission's high level commitment to open science, which continues to gain political momentum. Openness will be vital for data-driven science in Europe and for ensuring the European progress and prosperity in the future. This encompasses Open Innovation, Open Science and Open Access, and represents a paradigm shift in the way we do data science. It does not mean a change in the deep qualities of science but in the way we do it, and this change is fundamental to the future of our science innovation system including the whole research cycle and its stakeholders.

A recent European Commission public consultation in Europe (European Commission, 2015) shows that science is in transition to a more open mode of practice, which could be termed 'Open Science'. One current need is to identify the necessary policy-making adaptations to support this irreversible movement, which is driven by the availability of digital technologies, exponential growth of data, globalisation, emerging challenges, and the expectations of citizens and digital natives for accountable, responsive and transparent science.

This movement towards greater openness offers great opportunities for science in terms of value for money, more transparency, openness and networked collaboration, and more efficiency, reliability and responsiveness. However, key barriers to successful collaboration include data quality concerns, lack of credit for data work, lack of adequate infrastructures and limited awareness of the potential benefits of open data, as well as concerns about ethical and privacy issues. The explosion in the amount of data produced combined with advances in data science results in new opportunities for the use of big data in data-driven science. Enormous opportunities for science exist but, in order for data-driven science to take off, certain framework conditions need to be met, including data protection, copyright and research infrastructure modernisation.

In order to progress future open science policy, European Commission initiatives include supporting big data infrastructure, improving framework conditions, such as reducing barriers, providing incentives, and making science more efficient, reliable and responsive to societal needs. Currently, most open data initiatives exist within the individual Member States. There is a need for collaboration from the scientific community in supporting the move to a pan-EU model for data-driven science.

The European Open Science Policy Agenda will advance with a conference under the Dutch Presidency on open science planned in April 2016. Initiatives will include the establishment of a policy platform; the creation of incentives to promote best practices, research integrity and citizen science; and the removal of barriers, including the proposal of a European code of conduct and the promotion of mainstream open access to publications and data, the development of EU Guidelines on Intellectual Property Rights, funding for data management projects and the possible extension of the Horizon 2020 pilot on Open Research Data.

Open Research Data is part of the broader change towards more data-driven open science described herein. It refers specifically to making research data freely available for reuse beyond the purpose for which it was originally collected. This practice will aid further discovery and make scientific process more cost efficient and reliable.

The European Open Science Policy Agenda will also include the development of infrastructures to support open science, such as interfaces and data standards and funding. One key infrastructural element is a European Open Science Cloud for research data, protocols and methodologies. This is a part of Europe's ambition to support the transition to Open Science and make the most of data-driven science. Such a virtual environment for European researchers to store, manage, analyse and reuse data could be derived from a federation of existing and emerging data infrastructures. It would need to fulfil the strongly stated need for cost-effectiveness, privacy and awareness of intellectual property rights, as well as provide added value in terms of scale, data-driven science, interdisciplinarity and the ability to support the use of data to create knowledge and innovation. This major commitment on an EU-wide basis is foreseen over the coming years with initial elements being available from 2017.

1.2. The European Commission's Open Data Strategy and the EU Open Data Portal

The reuse of public sector data has steadily become more and more important in the political discourse and in practice over the last 15 years. However, barriers to embracing this trend are still present among data creators and owners and some concerns, including those related to privacy need to be carefully taken into account. Still, many of the reasons given for reluctance in data sharing, such as exposure of poor data quality, the likelihood of incorrect conclusions or the making of profit based on shared data, are not seen as valid by open data proponents. The fact remains that data created using public money should be freely available to citizens and companies who have a use for it and this is the fundamental principle of the EU Open Data policy.

Open Data policy is expected to foster economic activity, jobs and growth with direct economic benefits conservatively estimated at €40 billion per year. Further benefits include enhanced transparency, openness and civic participation, and better governance through data-based decision-making. Therefore, it is an important part of the 'data strand' of the Digital Single Market Strategy of the European Commission. Provision of the government food safety official controls data to a private application development company was presented as an example whereby citizens gain access to better information provided through a third party at no additional cost to the government above the creation of the data, which they are already doing.

The legal framework for open data sharing derives from Directive 2003/98/EC[1] on the reuse of public sector information (PSI) and was followed by a Decision on the reuse of the Commission's own resources in 2006 (amended in 2011 by Commission Decision 2011/833/EU)[2]. Directive 2013/37/EU[3] amending Directive 2003/98/EC reinforces the EU commitment to open data. This is in line with international trends in government policy including the G8 Open Data Charter adopted in 2013; the US Executive Order requiring government information to be open and in machine readable format by default; and the conclusions of the European Council in 2013 that 'Open data is an untapped resource with a huge potential for building stronger, more interconnected societies that better meet the needs of the citizens and allow innovation and prosperity to flourish. Interoperability and the re-use of public sector information shall be promoted actively'.

[1] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the reuse of public sector information.
[2] Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents.
[3] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the reuse of public sector information.

The EU PSI platform[4] encourages a community of PSI experts to collaborate and enhance the availability of data for reuse throughout the EU. A scoreboard of each country's performance is available via a 'crowdsourced' tool to measure the status of Open Data and PSI reuse throughout the EU including the open data community's activities.

The EU Open Data Portal[5] is a single point of access to the EU public sector data that brings together fragmented data sources from different EU institutions. It provides a harmonised way to describe metadata and improve data management, search capability in the form of human search interface and machine-to-machine search, and engagement with the data consumer community in the form of surveys, social media, mechanisms for feedback and dataset suggestions, as well as shared applications based on the EU data.

This portal includes all types of data from a wide variety of the EU institutions and agencies, not just scientific data. Users of the portal vary, including companies building new applications based on different types of data (17%), citizens who want to better understand what the EU institutions are doing with public money, and data-journalists, researchers and students.

Data provision through the portal has risen throughout the period from January 2013 to September 2015 and continues to do so in terms of both data providers and datasets provided. Data are provided by 56 data providers: the EU institutions, including the European Parliament, the Council, the European Investment Bank, the European Central Bank, the European Economic and Social Committee, the Committee of the Regions, Commission services, such as Eurostat, numerous Directorates General and a range of agencies. EFSA is soon to be added.

It is important to note that the current situation comprises over 150 data portals across the EU with different metadata and limited accessibility. While providing many useful datasets, this allows for limited interoperability capability. The pan-European Open Data portal, scheduled to launch (beta version) in November 2015, will provide a single gateway to public data from the EU Member States and a catalogue of information sources. Interoperability with other global portals will be important in the future but the EU must first ensure that we achieve what we need to do on interoperability issues, although discussions are ongoing with US and international partners.

In conclusion, the ongoing data revolution will affect all sectors of society and food safety regulation must embrace this change. Open data and open science have an enormous social and economic potential. In order to enable development of this potential, the Commission is practicing open data accessibility with their own data and aiming to support this practice by other EU data providers.

1.3. Data visualisations: drawing actionable insights from science and technology data

In an age of information overload, the ability to make sense of vast amounts of data and information and to generate insightful visualisations and knowledge from them is as important as the ability to read and write. Data visualisation is a powerful tool to help locate and understand the extent and power of our collective scholarly and technological knowledge. It provides a visual interface to large amounts of data from diverse sources from which potential clusters and trends can be identified, as well as emerging areas of research and scientific expertise. Data can be extracted from publications, patents, clinical trials, social media, etc., analysed and visualised, and subsequently used to generate new insights and knowledge.

The Indiana University Places & Spaces: Mapping Science exhibit comprises more than 100 maps of science and technology. Now in its 10th year, it introduces data visualisation science mapping techniques to the general public and to experts across diverse disciplines for educational and scientific purposes. The intention is to inspire cross-disciplinary discussion on how to track and communicate scholarly activity and scientific progress on a global scale.

Professor Börner illustrated, through examples and materials from the Information Visualisation Massive Open Online Course (MOOC)[6] at Indiana University, and a selection of maps from the Places and Spaces: Mapping Science Exhibit,[7] the wide application of data visualisation techniques that can be used to render data into insights. Examples included mapping and visualisation of the evolution of co-authorship and collaboration networks (Ke et al., 2004); the generation and subsequent citation of

[4] http://www.epsiplatform.eu
[5] http://open-data.europa.eu/en/data/
[6] http://ivmooc.cns.iu.edu
[7] http://scimaps.org

global scientific publications, the Global Scientific Food Web (Mazloumian et al., 2013); the impact of research & development investments on US innovation and global competitiveness (Council for Chemical Research, 2009); the distribution of language communities using social media (Fischer, 2011, 2012); and the impact of air travel on the global spread of infectious diseases (Colizza et al., 2006, 2007), as well as the identification of disease trends and alerts using real time data from the PREDICT health consortium (Brownstein et al., 2014). Figures 1–4 provide examples of selected maps.

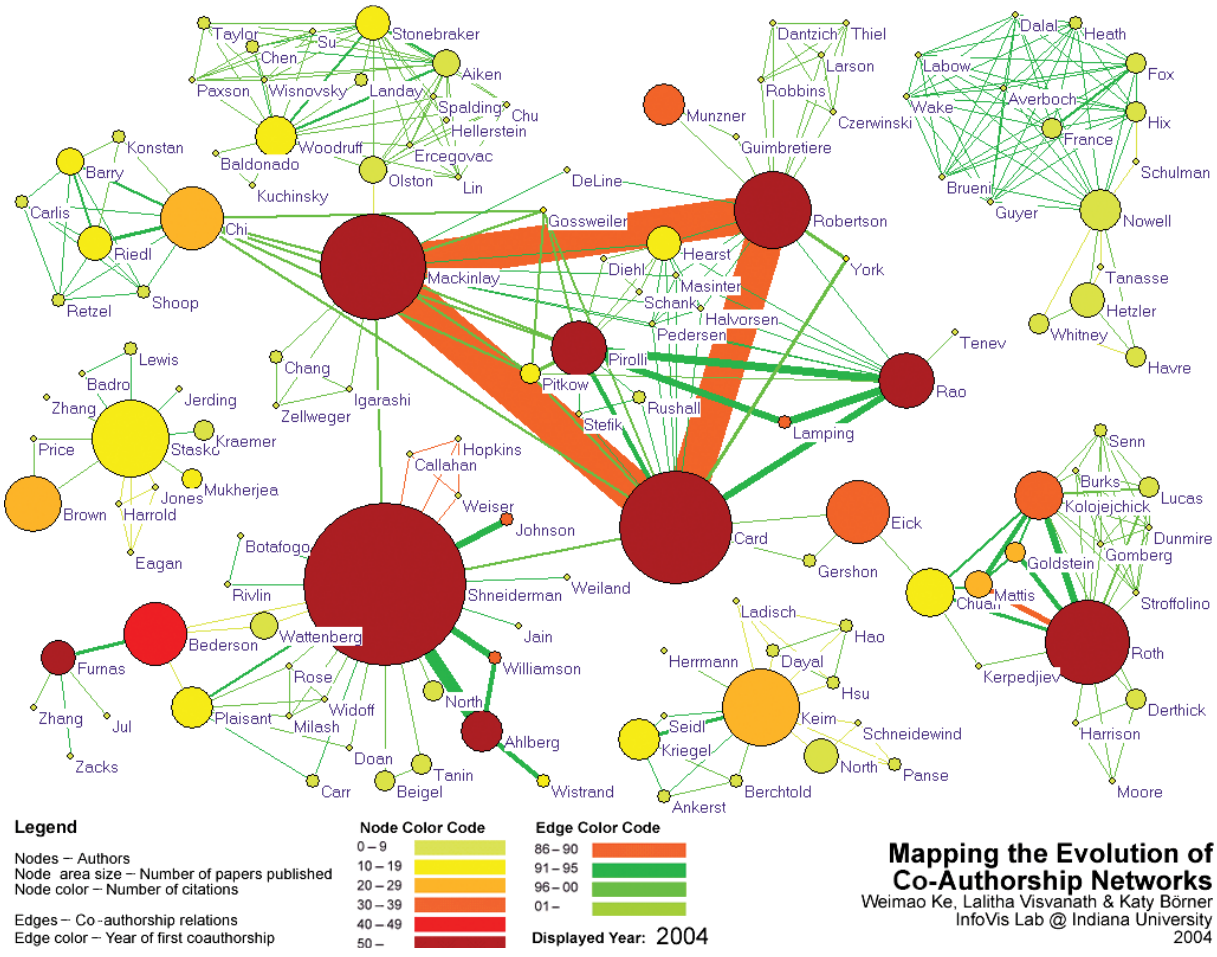

Figure 1: Co-authorship and collaboration networks (Ke et al., 2004)

Access to high quality and high coverage data is a precondition for generating high-quality data mining and data visualisation results. Much of the existing data visualisation research activities entail the use of freely available data and open code tools. The tools support the plug-and-play of different algorithms and the design of custom workflows; workflow logging makes it easy to rerun analyses. Examples include the *Science of Science* tool, which is funded by the US National Science Foundation and freely available.[8]

[8] https://dcf.efsa.europa.eu/dcf-war/dc

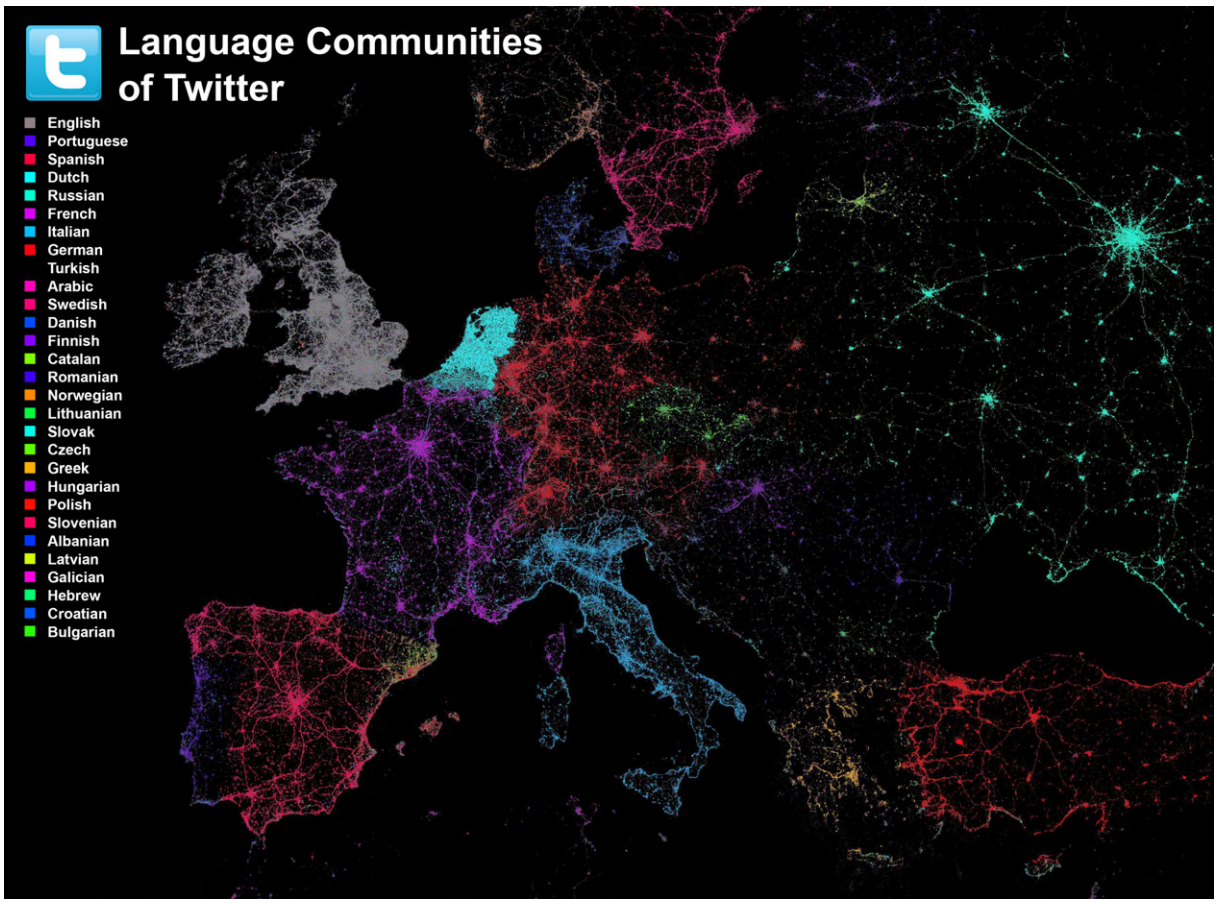

Figure 2: Language communities on Twitter (Fischer, 2011, 2012)

Although the application of data mining and visualisation tools offer enormous potential in the areas of science and technology, as well as food safety, a key challenge is to empower experts and the general public to generate and read data visualisations to produce reliable analyses. In this regard, 'open education', such as MOOCs (Bosman et al., 2016), free online courses that teach data visualisation skills, has an important role to play. Many companies have already started to embrace MOOCs to promote data visualisation literacy, and this trend is likely to continue.

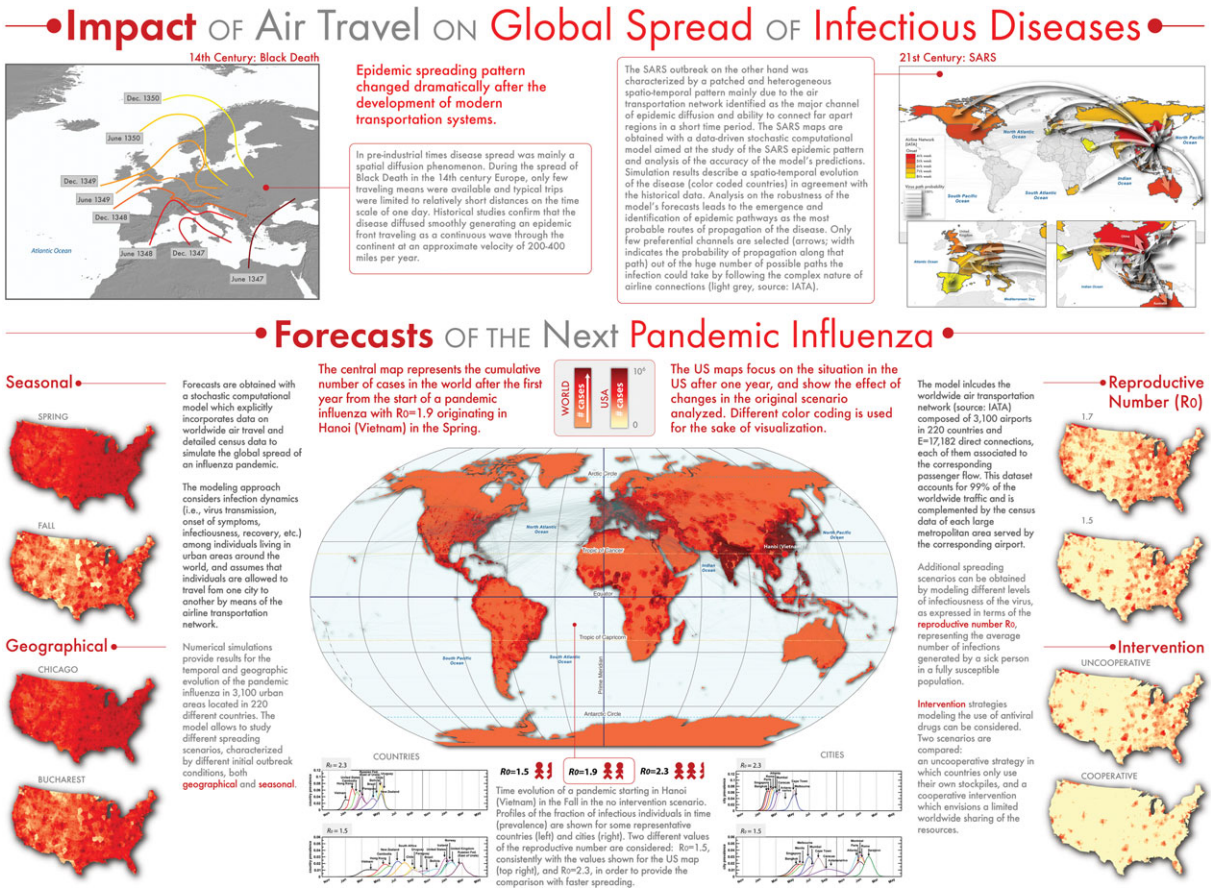

Figure 3: Global spread of infectious diseases (Colizza et al., 2006, 2007)

A second challenge is to develop an infrastructure where scientific papers and risk assessment outputs would allow access not only to underlying data, but also to workflows used to generate results. This would allow others to reproduce results, thereby increasing transparency, and to generate new knowledge by plugging in new algorithms and/or data into the workflows. Ideally, such infrastructures would become a standard part of the scientific work and review process.

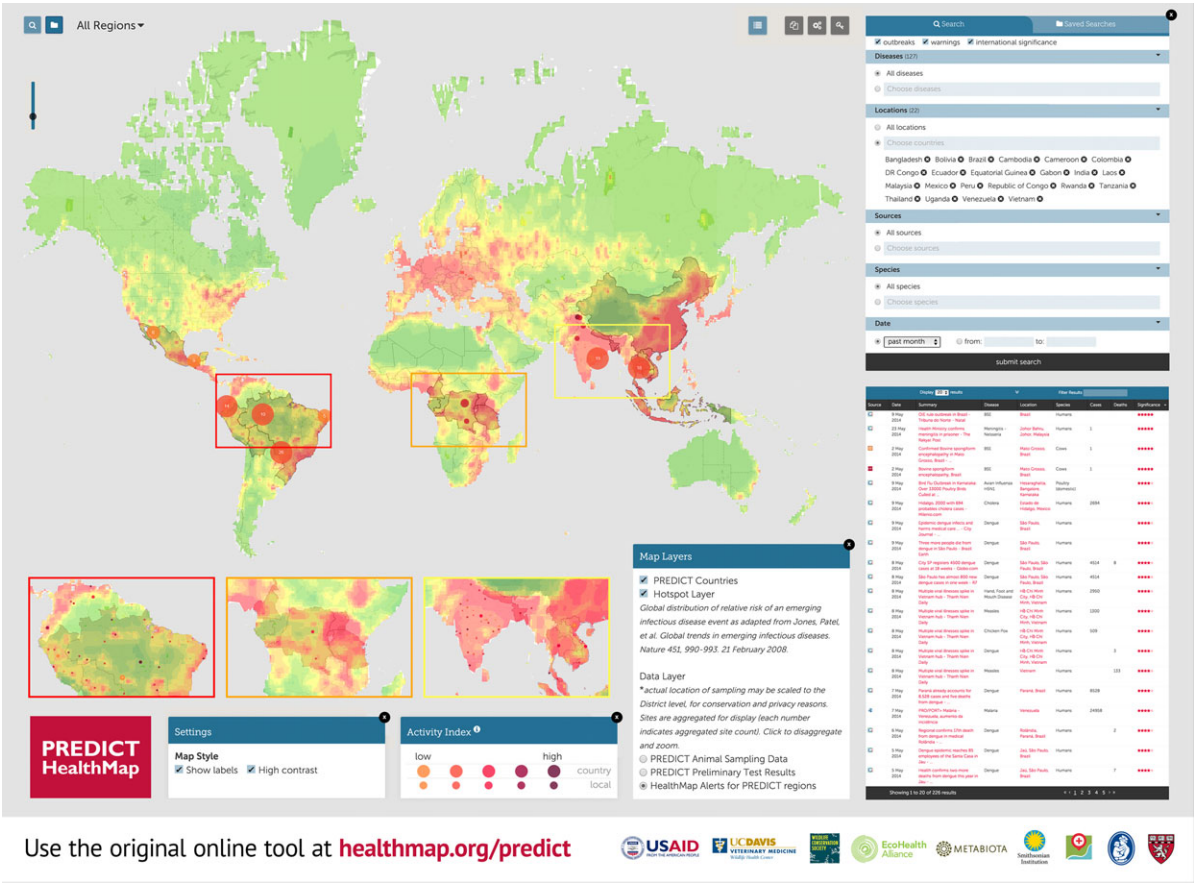

Figure 4: PREDICT health map (Brownstein et al., 2014)

1.4. Data interoperability and linked data technologies

Twenty-five years ago, the World Wide Web was invented and, subsequently, the data produced each day has massively grown not only in volume, but also in velocity. Every person, every company, and increasingly every activity, generates a huge stream of data. Yet, with a big challenge, our ability to make sense of that data, to turn that data into information, to make decisions based on that data has not accelerated at the same rate as the generation of data itself. Thus, some of the economic benefits around open data and our ability to use big data to solve some of the more pressing problems, including in the area of food safety, are still far from being realised.

The tweet 'David is checking in with the EFSA at the Stella Polare' can be clear for delegates present at the EFSA conference, but is more difficult for machines to interpret. 'EFSA' is the European Food Safety Authority and 'Stella Polare' is a conference centre in Milan. But EFSA could also be interpreted as the 'European Federation of Sea Anglers' or the 'Egyptian Financial Supervisory Authority' and 'Stella Polare' could be interpreted as a star as well as an Italian Navy Yacht. Thus, for a computer, David can be part of a fishing expedition organised by European Federation of Sea Anglers in an Italian Yacht. For machines, making sense of data is even harder than for humans.

Information identifiers are key to unlocking the huge amount of data in the open web and making sense of it. The foundation of the web is formed by connections that hold pieces of information together. A key value of open data is that it is allows others to reuse and generate new insights from this data. The full potential of open data is, however, restricted if it is not easily discoverable by others, including machines. In order to maximise the potential of open data, it is necessary to use open identifiers to explain what data is referring to.

Information identifiers are labels used to refer to an object being discussed or exchanged, such as products, companies, locations, ingredients or people. They allow the unambiguous and precise identification of information across different contexts and communities. If you are a consumer of data, open identifiers will allow you to connect more effectively information originated within and outside

your organisation or institution. Likewise, if you are a producer of data, using information identifiers, you make sure that data is more valuable and easier to integrate across different sources. If you publish those identifiers and make them 'open', it will extend the value of your identifiers across a much broader community.

An identifier is not just primary key in a closed homogenous database system. An identifier is a way of uniquely, unambiguously, precisely identifying a thing or object in the vast, broad, open, heterogeneous environment that is the World Wide Web. In this complex context, an identifier scheme is the framework needed to provide rules for creating good identifiers, making sure that are both human and machine readable.

For many years, Thomson Reuters addressed internally the problem of unequivocally identifying things and assigning unique identifiers. In 2015, Thomson Reuters launched 'Open Perm ID' where part of those Identifier Framework Schemes developed by Thomson Reuters was made available to the public for free either via a web-interface or via machine application program interfaces (APIs).

Open Perm ID provides three main services:

- searching for an object to retrieve its machine readable identifier (i.e. Perm ID);
- submitting structured data to match all records, resolve their content and retrieve their PermIDs; and
- submitting unstructured data (e.g. a text, a blog, a research report) to extract and identify different entities and return their PermIDs.

The last one, the intelligent tagging service, also creates links and relationships between entities. With these links and relationships a 'knowledge graph' can be created to allow finding additional connections and facts around the data that would not be identified using normal searches or data mining techniques.

The Open Data Institute in the UK and Thomson Reuters wrote a joint white paper on the added value of identifiers for open data (Open Data Institute & Thomson Reuters, 2014). The paper proposes some guiding principles, both for those creating and managing identifier schemes and those who are using them, and examines why and how information identity must evolve from being an inherent part of dataset design to being a distinct discipline in its own right.

In conclusion, humans are losing the 'data race', but is it the same for machines? Can machines, such as IBM Watson, close the gap and make sense of all the data produced daily? Thomson Reuters recently announced collaboration with IBM Watson. However, even with cognitive computing technology, there is still the need to build a common reference underlying the data. In this, people are better than machines: you need an authority to solve master data issues, to perform data curation or to define a 'social' framework to describe how data are used, what it means and how it relates across different communities. We always need the combination of machines and people to make the sense of data at best.

Perhaps the technology used by Thomson Reuters is not unique. What is unique is the way that Thomson Reuters have applied it, placing the consumers of data at the centre. The publication of the white paper around Information Identifier Schemes, which promotes the use of standards, as well as the publication of the Thomson Reuters ontologies and taxonomies, is unique.

1.5. Opening clinical trial data

During the past years, Industry and Regulators, including the European Medicines Agency (EMA), faced several requests to make clinical trial data public. This was a saga that unfolded over almost a decade and started out on mistrust, misjudgement and misalignment, but ultimately came to a very positive happy ending.

Medicines are among the most regulated products. A marketing authorisation is needed to sell a medicine in a country and this authorisation is obtained by demonstrating the Quality, Safety and Efficacy (QSE) of the product by submitting a dossier to a Regulator, which, in Europe, is usually the EMA. A dossier consists of a pharmaceutical part (describing the product's stability, sterility, etc.), a non-clinical part (reporting the animal studies results) and a clinical part (reporting the results on (randomised) clinical trials in patients). Clearly, the latter is the most sensitive, most expensive to generate and most valuable part of a dossier for Industry, who want to share this information only with the Regulators granting the market authorisation.

Academics pushed, first Industry and then Regulators, in making this information public, their argument being: data on which a marketing authorisation is based should be public. Industry and,

initially, the EMA refused to disclose clinical trial data under the argument that it must be considered commercially confidential. The discussion was at times confrontational. After many years of discussions and some court cases, it became clear that the question was no longer whether Clinical Trial documentation would be published or not but which parts of the dossier should be published.

In 2012, the EMA decided that, in future, for every new dossier application received, clinical trial data would go public after the end of the marketing authorisation process. The EMA invited all concerned stakeholders to participate in the preparation of a policy to define the details about the open access. After consultation, the policy was adopted (EMA, 2014), finding an appropriate balance between the different competing interests: protecting commercial confidentiality and patient privacy, avoiding unfair commercial use and, at the same time, being open and transparent. The policy foresees two levels of access: one without the need for registration, where data can only be seen on the screen but not be downloaded or printed, and another level with the need for registration, where the information can be printed or downloaded, subject to the signature of a terms of agreement on the intent to use the data.

Looking at the statistics about the requests to access clinical trial data, even before the policy entered into force, the number one requestor is Industry itself, which clearly wants a competitor's data so that they can learn from each other's experiences. Making clinical trial data public will improve the efficiency of clinical trials and, ultimately, the medicines development process. It will enable the creation of quality-controlled databases informing future research questions and projects, and allow better analysis of patients' data from different trials avoiding silo-limited analyses.

The train has left the station: the EMA has stimulated the pharmaceutical Industry to be proactive in sharing clinical trial data, moving the debate from 'all data is confidential' to 'what is confidential in clinical trial reports' and 'how to protect those data from unfair commercial use'. We have moved from a lack of trust to constructive use of data supporting future medicines development.

In the future, clinical trial data will be available for secondary research purposes. Already, some organisations have become custodians of Industry's data and, on behalf of them, they administer and allow access to Industry's data behind firewalls. Accessing, merging and analysing multiple companies' data in one single place is the future. Following the adoption of the EMA's policy, the risk of legal action against the EMA is not totally eliminated, especially in cases where the terms of agreement about the use of disclosed data are violated. To minimise this risk, the EMA has paid particular attention to writing the terms of agreement.

1.6. Data collection by public bodies: Joint EFSA–Member State initiatives

EFSA's core business is the produce scientific opinions on risk assessments of foodborne hazards at the EU level. Food safety risk assessments comprise four steps: hazard identification, hazard characterisation, exposure assessment and risk characterisation, each of which requires access to data. Hazard identification and hazard characterisation typically rely on toxicology data from animal, biological and/or human studies. Dietary exposure assessments require data on the occurrence and levels of microbiological or chemical hazards in food, as well as food consumption data. A food classification system (e.g. FoodEx food classification and description system; EFSA, 2015a) is needed to describe and categorise food in order to link data from occurrence and food consumption databases with estimated dietary exposure. The last step, risk characterisation, combines data from hazard characterisation and exposure assessment to estimate a likelihood of risk associated with a given exposure.

EFSA relies on data from the Member State laboratories to estimate dietary exposure. The Evidence Management (DATA) unit of EFSA is responsible for collating occurrence and food consumption data from the EU Member States. It receives support for its data collection activities from EFSA's advisory forum, the EFSA focal point network, EFSA scientific networks on data collection (comprised of data provider representatives from national competent authorities) and industry stakeholders. Good collaboration between EFSA and the Member States is crucial in this regard.

The majority of foodborne hazard occurrence data sent to EFSA comes from laboratories involved in national monitoring programmes via national competent authorities. Other data providers include the food industry (mainly via industry associations), universities, consumer associations and, in some cases, the European Commission (e.g. Joint Research Centre).

Given that EFSA receives an increasing volume of data from the Member States on an annual basis, standardisation and harmonisation activities are essential to ensure efficient and effective data

transmission to support EFSA's needs. EFSA developed the Standard Sample Description (SSD; EFSA, 2010, 2013), with support from the Member States, which is the EFSA standard for transmission of foodborne hazard analytical occurrence data from the Member State data providers to EFSA. The SSD data model contains approximately 80 standardised data elements (fields) describing metadata (e.g. analytical method, date of sample analysed, etc.) associated with an analytical result, of which approximately 20 are mandatory. Following the definition of the SSD standard, the Member States invested a lot of resources and efforts in collecting and harmonising data. In the case of pesticide residues, reporting of residue monitoring data from the EU coordinated multiannual control programmes in SSD format to EFSA is mandatory (Commission Implementing Regulation (EU) No 788/2012)[9]. The added value of the SSD is that occurrence data from several food safety domains (e.g. chemical and biological hazards) can be reported to EFSA in a standardised format, thus increasing efficiency and bridging silos between data domains at national level. Other benefits include improvements in data quality, as well as harmonisation of data from national control and monitoring activities to support risk assessment activities at the national level.

When managing large volumes of data, the use of standard terminologies (information identifiers) is necessary. A standard terminology for food is particularly important with respect to dietary exposure assessments.

Data quality is of paramount importance. EFSA provides tools i.e. the Data Collection Framework (DCF) web interface[8] to receive and validate data received from the Member State data providers. The DCF contains in-built business (validation) rules to perform automatic checks on incoming data. Automatic feedback is provided to the data provider; if a data file is rejected, the data provider subsequently corrects the data file and resubmits through the DCF. These automatic checks serve as a first quality control on incoming data. Additional quality checks are undertaken by EFSA data managers. Finally, during the data analysis stage, clarification requests on a dataset may be sent to data providers by EFSA data analysts. At all stages of the data checking process, data providers correct the data, if necessary, and resubmit it to EFSA.

EFSA acts as a central storage hub of European risk assessment data and provides data visualisation tools for extraction and analysis. Storing national data in a central repository maintained by EFSA provides additional advantages. Of special interest is the possibility to use up-to-date systems and data for exposure calculations at national level. EFSA recently developed a Scientific Data Warehouse (sDWH), which currently contains data on food consumption, chemicals (contaminants, food additives and pesticide residues), zoonoses and chemical hazards (e.g. toxicology data extracted from EFSA opinions). The sDWH provides an interface to access EFSA's data hub. The goal is to increase transparency of EFSA's work and, ultimately, to allow new knowledge to be created through reuse of accessible data. The EFSA's sDWH access rules outline different levels of access to data for different stakeholders (EFSA, 2015b).

In an effort to promote a more open approach to data and to foster stronger collaboration and data sharing between the Member States, EFSA recently initiated a pilot study – the 'Circle of Trust' initiative – with interested Member States within the framework of the EFSA scientific network on chemical occurrence data.[10] Members of the 'Circle of Trust' agreed to allow access (to other members of the 'Circle of Trust') to contaminant occurrence data submitted to EFSA from monitoring programmes, under agreed conditions, using the sDWH. The pilot will continue during 2016.

EFSA's data collection and harmonisation activities with the Member States are relevant to the entire infrastructure of data collection at national level, and not solely to the national competent authorities which are EFSA's primary national interfaces. In many cases, the national competent authorities act as national data hubs for data generated/collected by local authorities and laboratories. Thus, in some cases, data arrives at EFSA via a chain of data providers at the national level. This needs to be considered by EFSA when implementing changes and/or updates to data collection and transmission models, processes and tools. Although the majority of national competent authorities are familiar with EFSA's systems for data collection, validation and extraction via the sDWH, EFSA and the

[9] Council Regulation (EC) No. 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No. 847/96, (EC) No. 2371/2002, (EC) No. 811/2004, (EC) No. 768/2005, (EC) No. 2115/2005, (EC) No. 2166/2005, (EC) No. 388/2006, (EC) No. 509/2007, (EC) No. 676/2007, (EC) No. 1098/2007, (EC) No. 1300/2008, (EC) No. 1342/2008 and repealing Regulations (EEC) No. 2847/93, (EC) No. 1627/94 and (EC) No. 1966/2006.

[10] Commission Implementing Regulation (EU) No. 1337/2013 of 13 December 2013 laying down rules for the application of Regulation (EU) No. 1169/2011 of the European Parliament and of the Council as regards the indication of the country of origin or place of provenance for fresh, chilled and frozen meat of swine, sheep, goats and poultry.

national competent authorities have a joint responsibility to ensure that this knowledge is cascaded to national laboratories responsible for generating the data. This will lead to continued improvements in data quality and also to increased efficiency at all steps of the data life cycle.

EFSA and the Member States have made considerable progress in the area of data collection, data validation and standardisation through the cocreation of data transmission standards, platforms and data exchange tools. However, much remains to be done in particular given EFSA's vision to adopt a more open approach to data. This will entail more dialogue and engagement with a wider range of national stakeholders including industry. In this regard, data governance will become an important issue to address. There is a clear need for better clarity on data governance at the EU level. This is one of EFSA's and the Member States main challenges to address with respect to open data.

There are also discussions at a Member State level on how to revise national priorities and, in the long run, to support to the European project. A broader engagement in these discussions involving, for example, more risk managers and politicians would facilitate the future development of even more efficient data collection systems to support more transparent risk assessment of foods on the European market.

1.7. METRO GROUP's Global Standard Traceability Solution

Transparency and traceability are central to gaining trust throughout the global supply chain. To this end, industry has a key responsibility to ensure robust product traceability. In order to achieve this in an efficient way, there is a need to easily access and connect data from several actors along the food supply chain.

METRO GROUP/MAKRO Cash & Carry, one of the METRO GROUP sales lines, which deals with wholesale trade, is one of the largest fish retailers in Europe, represented in 28 countries globally. In order to provide robust and scalable product traceability on an international level, the company has established, in 2013, a Global Standard Traceability Solution (GSTS), with an initial focus on fresh fish. GSTS is a decentralised technology solution applicable to food and non-food products that links information on the journey of products from source to the retail customer and, ultimately, to the final consumer. The system is based on global standards provided by GS1 (standards organisation) to ensure global interoperability and best use of modern technology. Its innovative aspect is that – from producer to retailer – all parties can enter their product details in a central database or in their own databases, which are subsequently linked through a unique product identification code. Via this code, it is possible to retrieve information about particular products along their supply chain.

When purchasing fresh fish, for example, customers of METRO's wholesale subsidiaries can use a smart phone app to find out in which waters and with which method the fish was caught. The customer scans the bar code of the fresh fish using METRO's *Pro Trace* app; the bar code contains a lot number (identifying a particular lot) and a GTIN (Global Trade Item Number) which links the fish to product information uploaded by suppliers in decentralised databases on a cloud (e.g. date of catch, fishing vessel, scientific name, location of catch, certificates, etc.). All information along the supply chain is linked to the product via the unique product identifier code.

In addition to increasing transparency for the customer and, ultimately, the end consumer regarding origin, composition and life cycle of a product, the solution serves to optimise internal processes during product recalls; in the event of a potential food fraud scandal, it can also help to identify the potential source of fraud and the production of the product through the supply chain. It also fulfils METRO's legal obligation at the European level and the international level regarding traceability (Council Regulation (EC) No. 1224/2009[9]; Commission Implementing Regulation (EU) No. 931/2011[11]; Commission Implementing Regulation (EU) No. 1337/2013[10]), and it reduces supply chain complexity.

The aim is to extend the solution to a wider geographical coverage and product base (e.g. food of animal origin, fresh fruit and vegetables, as well as non-food products). The ultimate goal is to strive for one global solution based on global harmonised standards that cannot be achieved without a joint collaborative effort of several stakeholders.

METRO's GSTS illustrates the value of data accessibility, data interoperability (connectivity between information sources) and common global standards to implement food product traceability on a global scale. The solution is open and adaptable allowing further product categories or additional databases

[11] Commission Implementing Regulation (EU) No. 931/2011 of 19 September 2011 on the traceability requirements set by Regulation (EC) No. 178/2002 of the European Parliament and of the Council for food of animal origin.

to be added and is open to all market participants along the supply chain worldwide. A lot can be learned from experience gained and customer feedback to date which, through further dialogue and collaboration, could be applied across a wider food safety remit.

2. Conclusions

We are entering a new era of openness with the advent of open data. There is an irreversible paradigm shift in the way all sectors within the food safety risk assessment remit and beyond deal with data. Open data has enormous social and economic potential, in particular, for data rich public institutions, such as EFSA: more transparency, more innovation, more possibilities for global networking, a higher standard of science and clinical trials, as well as better decision-making. The plethora of benefits associated with open data outweighs its perceived risks (e.g. potential misuse/ misinterpretation of data).

Our ability to turn data into information and to make decisions from data has not accelerated to the same degree as the emergence of open data. Information identity (use of identifiers) is a key enabler to generate added value and new knowledge from open data.

Increasing openness and transparency will continue to dominate the future of food safety risk assessments. Data hugging will become a practice of the past. EFSA and its data providers need to embrace this paradigm shift. Within this context, the following core needs were identified as an outcome of the breakout session discussions:

- Align strategies between the European institutions and establish early dialogue and stronger collaboration at the European and international level to improve interoperability between datasets and use of common data standards.
- Broader multisector, interinstitutional and interagency engagement to reap the value of prior knowledge, experience and tools developed by other sectors and organisations, and to avoid duplication of efforts (thereby wasting resources).
- Promotion of a culture of open science by including as a precondition for funding allocation provision of data, data management plans and data analysis code in funding applications; in addition, publishers have an important role to play by requesting that underlying data, code and workflows accompany the publication of scientific outputs.
- Development of expertise and empowerment of (young) scientists, researchers and the general public to generate new knowledge and insights from open data, as well as associated data analysis and visualisation tools.
- Establishment of clear data governance at the EU level within the food safety risk assessment remit.
- Establishment of collaborative processes across the EU to ensure effective investment in infrastructure, capacity and capability to leverage the potential of open data in food safety.

References

Bosman A, Brent P, Cocconcelli PS, Conole G, Gombert D, Hensel A, Kardassi D, Kneifel W, Koulouris S, Soares AMVM and Zilliacus J, 2016. Special issue: expertise for the future: learning and training in the area of food safety risk assessment. EFSA Journal 2016;14(S1):s0503, 10 pp. doi:10.2903/j.efsa.2016.s0503

Brownstein J, Joly D, Karesh W, Daszak P, Wolfe N, Goldstein T, Aman S, Freifeld C, Mekaru S, O'Rourke T, Morse S, Kreuder Johnson C, Mazet J and the PREDICT Consortium, 2014. PREDICT: HealthMap. Courtesy of USAID (US Agency for International Development), University of California, Davis, School of Veterinary Medicine, and Children's Hospital Boston. In '10th Iteration: The Future of Science Mapping', Places & Spaces: Mapping Science. Edited by Katy Börner and Samuel Mills. Available online: http://scimaps.org

Colizza V, Barrat A, Barthélemy M and Vespignani A, 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. Proceedings of the National Academy of Sciences USA, 103, 2015–2020.

Colizza V, Vespignani A and Hardy EF, 2007. Impact of air travel on global spread of infectious diseases. Courtesy of Indiana University. In: Börner Kand Davis JM (eds.). *3rd Iteration (2007): The Power of Forecasts', Places and Spaces: Mapping Science*. Available online: http://scimaps.org

Council for Chemical Research, 2009. Chemical R&D powers the U.S. innovation engine. In: Börner K and Hardy EF (eds.). *'5th Iteration: Science Maps for Science Policy-Makers' Places & Spaces: Mapping Science*. Available from: http://scimaps.org

EFSA (European Food Safety Authority), 2010. Standard sample description for food and feed. EFSA Journal 2010;8(1):1457, 54 pp. doi:10.2903/j.efsa.2010.1457

EFSA (European Food Safety Authority), 2013. Standard sample description version 2.0. EFSA Journal 2013; 11(10):3424, 114 pp. doi:10.2903/j.efsa.2013.3424

EFSA (European Food Safety Authority), 2015a. The food classification and description system FoodEx2 (revision 2). EFSA supporting publication 2015:EN-804, 90 pp. Available online: http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/804e.pdf

EFSA (European Food Safety Authority), 2015b. The EFSA data warehouse access rules. EFSA supporting publication 2015:EN-768, 18 pp. Available online: http://www.efsa.europa.eu/?sites/default/files/scientific_output/files/main_documents/768e.pdf

EMA (European Medicines Agency), 2014. Policy on publication of clinical data for medicinal products for human use (EMA/240810/2013). Available online: http://www.ema.europa.eu/?docs/en_GB/document_library/Other/2014/10/WC500174796.pdf

European Commission, 2015. Validation of the results of a public consultation on science 2.0: science in transition. Directorate General (DG) Research and Innovation. Available online: http://ec.europa.eu/digital-agenda/en/news/final-report-science-20-public-consultation

Fischer E, 2011. Language communities of Twitter. Flickr. Last modified October 24th 2011. Available online: http://www.flickr.com/photos/walkingsf/6277163176/in/photostream/

Fischer E, 2012. Language communities of Twitter. Courtesy of Eric Fischer. In: Börner K and Stamper MJ (eds.). *'8th Iteration: Science Maps for Kids', Places & Spaces: Mapping Science*. Available online: http://scimaps.org

Ke W, Börner K and Visvanath L, 2004. Major information visualization authors, papers and topics in the ACM library. Analysis and Visualization of the IV 2004 Contest Dataset. Proceedings of the IEEE Symposium on Information Visualization, IEEE Computer Society Washington, DC, USA, page 216.1. Available online: http://cns.iu.edu/docs/publications/2004-ke-infovis.pdf

Mazloumian A, Helbing D, Lozano S, Light R and Börner K, 2013. Global multi-level analysis of the scientific food web. Scientific Reports, 3, 1167. doi: 10.1038/srep01167

Open Data Institute & Thomson Reuters, 2014. Creating value with identifiers in an open data world. Available online: http://site.thomsonreuters.com/corporate/pdf/creating-value-with-identifiers-in-an-open-data-world.pdf

Abbreviations

AMR	antimicrobial resistance
API	application program interface
DCF	Data Collection Framework
EMA	European Medicines Agency
GSTS	Global Standard Traceability Solution
GTIN	Global Trade Item Number
MOOC	Massive Open Online Course
PSI	public sector information
QSE	Quality, Safety and Efficacy
sDWH	Scientific Data Warehouse
SSD	Standard Sample Description