



Emerging Data Governance Issues in Big Data Applications for Food Safety

Salvatore Sapienza and Monica Palmirani^(✉)

University of Bologna, CIRSFID, Via Galliera 3, 40121 Bologna, Italy
{salvatore.sapienza2, monica.palmirani}@unibo.it

Abstract. The food industry and food safety authorities show an increasing interest in Big Data applications. On the one hand, Big Data strengthens data storage, data mashup, and methodology of risk assessment; on the other hand, the presence of risks and challenges linked to Big Data demands for future-proof solutions. While business practices and academic literature have already identified ways of exploiting the value of large collections of data, emerging governance issues including data sovereignty, competition, privacy and data donations, call for further analysis.

Keywords: Big Data · Food safety · Data governance · Emerging issues
Risk assessment

1 Introduction

As with other research areas, current trends in the agri-food domain highlight a growing interest in the collection and the use of Big Data. As the FAO's Sustainable Development Goals point out, the "Data Revolution" is reshaping the way in which policymakers interact with the agri-food industry, biotechnology research, and food security when implementing policies towards sustainability [1]. Among the areas positively affected by Big Data, food safety plays a crucial role in achieving FAO's commitment to promote the human right to qualitatively adequate food [2]. Therefore, Big Data is likely to represent a gateway to further develop a data-driven approach to food safety risk assessments [3].

This study departs from two premises. First, while the positive effects of Big Data can be achieved only through the collaboration of multiple stakeholders involved in the food safety domain, conflicting interests pertaining to the protection of Big Data as a valuable commercial asset hinder the finding of shared data governance. Secondly, although food consumption data is needed to understand how individuals and food interact, this information may reveal health status, thus requiring due attention to the safeguard of fundamental rights, including appropriate security measures. In this paper, three intertwined lines for further research are identified: first, the one related to the sovereignty and the exchange of information between different jurisdictions; secondly, the area related to the interferences between data ownership and the exchange of information performed by the corporations involved in this domain; lastly, the sphere connected to the cooperation between individuals and food safety assessors.

Following this introduction, Sect. 2 describes the research question and the research methodology. While Sect. 3 summarises modern uses of large datasets in food safety, Sect. 4 introduces cutting-edge technologies whose adoption by safety assessors is reasonably foreseeable. Section 5 illustrates the data governance issues that emerge from the discussion among food safety authorities, their stakeholders, and other interested parties. Lastly, final considerations summarise the findings of the previous sections and make recommendations for future work.

2 Research Questions and Methodology

Several clashing interests emerge from the use of Big Data for food safety purposes, thus originating specific issues whose impact has not been fully analysed. While the outcome of some technologies—including genetically modified organisms and nanotechnology—is currently under the scrutiny of prominent scholars, this short study raises open questions and possible lines of research to identify emerging and undetected data governance issues.

Big Data applications currently employed by risk assessors are grounded on the practices developed within corporations and academia. Therefore, this paper has been drafted following a technology-driven approach that illustrates the connections between food safety authorities, the food industry, other industry sectors, and the academic community. This research methodology has been developed by selecting two kinds of innovative practices.

On the one hand, the selected Big Data applications rely upon or produce large and structured quantities of data gathered from multiple sources. In this domain, heterogeneous datasets are vital to reduce the presence of gaps in data and bias in data model design and to achieve a higher level of precision in the risk assessment. On the other hand, Big Data applications under scrutiny require the joint efforts of a plurality of actors to exploit the potential of this data. Since data originates from multiple sources, a certain degree of collaboration is needed to appreciate the benefits of Big Data. Conversely, those applications which neither compute large quantities of data, nor involve multiple sources and entities, are out of the scope of this paper.

The need of future-proof solutions raises the urgency of discussing both traditional Big Data practices already employed by risk assessors and possible future applications whose adoption is likely to occur in the near future. Accordingly, the following sections categorise the selected Big Data techniques in line with this approach.

3 Current Big Data Applications for Food Safety

Current research has already identified several ongoing applications of Big Data in food safety [4]. For the purposes of this paper, “Big Data” will refer to the use of collections of data characterised by Volume, Velocity, Variety and Veracity used to extract Value¹.

¹ <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>.

This paper adopts an economically-neutral definition of Value. In other words, it should be understood as any advantage in the food safety risk assessment that can result in an increase in speed, precision, forecast reliability, and cost.

3.1 Data Analytics in the Food Safety Domain

Despite its business nature, data warehousing is a consolidated trend in the context of food safety. World Health Organisation's (WHO) "FOSCOLLAB" project enables users to access food safety data and information from multiple existing sources: Joint FAO/WHO Expert Committee on Food Additives (JECFA), Joint FAO/WHO Meeting on Pesticide Residues (JMPR), and WHO Collaborating Centres Database. FOSCOLLAB and its related databases are easily accessible through a digital platform. The European Food Safety Authority (EFSA) Data Warehouse has been active since 2015. Data related to zoonotic diseases, antimicrobial resistance, foodborne outbreaks, pesticide residues, chemical contaminants, and chemical hazards are accessible through an online access point. The platform has been developed to strengthen scientific progress by granting access both to the general public and food safety professionals, including EFSA's stakeholders and researchers. Similarly, the EFSA Comprehensive European Food Consumption Database collects data pertaining to food consumption in the EU. As regards the United States, the Food & Drug Administration (FDA) has opted for a different approach by making the data warehouse available to the public through its APIs. Since this accessibility method can be cumbersome to the general public, FDA web pages contain links to third parties' platforms. However, the number of apps pertaining to food safety seems quite limited.

This relevance of this large amount of information available for analysis is illustrated by Memorandum of Understanding signed by the EU European Chemicals Agency (ECHA) and EFSA. It reinforces their commitment of "[e]xploring the application of modern technology (e.g. artificial intelligence, machine learning and data mining)" which rely upon Big Data to spread their positive effects [4]. As the Open Data Institute has pointed out, data gaps and bias in the design of data analysis models constitute potential concerns when the aforementioned techniques are employed [5].

3.2 GIS-Based Approaches to Food Outbreaks Analysis

Geographic Information Systems' (GISs) practical implementations in food safety have been widely researched. The combination of geospatial data and analyses conducted on plants, foods, and feed has proven to be helpful in tracing the origins and preventing the spread of diseases.

A sample of the studies conducted using GIS-based approaches confirms their potential applications in identifying vulnerabilities in the food supply chain [6], forecasting contamination of crops on a regional basis [7], and identifying correlations between *Escherichia coli* O157:H7 and vegetable production [8]. Food safety authorities are willing to adopt GIS-based assessment methodology. The GIS-iRisk project results from the joint efforts of the FDA and NASA. It correlates GISs with predictive risk-assessment models to forecast when, where, and under what conditions risks for human health may emerge from crop contamination. Similarly, EFSA's

PERSAM software has been developed to predict the environmental concentration of pesticides in soil, in accordance with the guidelines adopted by the experts of the EFSA Pesticide Panel [9].

3.3 Whole Genome Sequencing for Foodborne Pathogens Identification

Whole Genome Sequencing (WGS) is rapidly reshaping the process of identifying and characterising foodborne pathogens. While previous techniques used to analyse only the biological components of the molecule under scrutiny, WGS is a universal methodology that can be appropriate for monitoring human/animal health and food [10].

Food safety authorities in Western countries are prioritising the adoption of WGS. The UK Food Safety Authority Chief Scientific Adviser's Science Report describes a successful case of WGS application in countering the Salmonella outbreak in 2014. WGS methodology corroborated the previous research by confirming the genetic link between UK cases and German eggs, thus identifying the source of the outbreak [11]. The US FDA is leading this paradigm shift by implementing GenomeTrakr Network, a decentralised network of laboratories whose main aim is to identify pathogens through WGS. GenomeTrakr Network consists of sixty-three private and public facilities, placed inside and outside the US. In June 2014, EFSA adopted a positive resolution on the immediate implementation of the WGS of foodborne pathogens for the protection of public health. EFSA stressed that, EU private and public entities involved in healthcare should strengthen their collaboration with international counterparts to harmonise the use of WGS [12].

4 Future Big Data Applications for Food Safety

This section outlines innovative ideas for future applications of Big Data within the food safety domain. Even though some proofs-of-concept have already been published, their full deployment requires time and effort and should thus be considered a long-term goal. These examples contribute to identifying possible emerging issues in data governance, therefore promoting the adoption of future-proof solutions.

4.1 Automated Text Analysis of Scientific Opinions

Existing literature demonstrates the advantages of analysing textual digital information using both linguistic and statistical methodologies for food safety purposes. Both news related to foodborne outbreaks and academic research on pathogens are the most observed sources of information. On the one hand, IBM FoodSIS adopts a machine learning algorithm based on a supervised ranking system that retrieves and classifies food incidents news on behalf of the National Environment Agency of Singapore [13]. On the other hand, text mining may be used in the identification of emerging chemical and biological risks from previous studies published in academic journals [14]. To the author's knowledge, the potential of automated text analysis in risk assessment documents published by food safety authorities - including the open-access and free of charge EFSA Journal, i.e. the repository of EFSA's scientific opinions - is largely

undiscovered. For instance, potential outcomes may consist of predicting trends in the use of chemical substance by analysing their recurrence in EFSA's opinion, identifying argumentative patterns in these documents, and simplifying scientific concepts to improve readability for the general public.

4.2 IoT, Smartphones and Social Media for Real-Time Food Consumption and Food Alerts

The number of food-related features in Internet-of-Things (IoT) devices is increasing. In 2017, EFSA launched a tender to explore 'collaborative models' for data gathering², thus confirming its interest in exploring decentralised methods of data collection which may include some of the devices discussed in this section.

Through third-party softwares, Amazon Echo suggests recipes for the user to cook, and as a result, gathers real-time food consumption data. This device is also widely used to manage users' grocery lists, thus helping to predict food consumption in the near future with a high level of certainty.

As with IoT devices, smartphone applications may be used to analyse consumers' behaviour. The FoodProfiler (formerly, Food Intake) is a mobile app developed by Wageningen University & Research to collect food consumption data directly from users [15]. User profiling allows to compare clusters of participants, to identify product combinations, and to analyse consumption trends. Positive outcomes may be expected from this research methodology since it combines the use of handy devices, a short recall time, and low burdens for the participant.

Lastly, a small body of literature has discussed the use of social media as food-related communication tools both in regular times [16] and in the course of emergencies [17]. Studies have confirmed the possibility of retrieving overweight and diabetes rates from a large corpus of tweets containing food-related hashtags [18]. New paradigms in gathering consumption data may consist of automatic identification of food in photos uploaded to social media. MIT's artificial intelligence algorithm Pic2Recipe! is able to identify the presence of food in pictures and retrieve the recipe³.

While the potential of this data to predict consumers' attitudes is currently exploited only by private entities, some beneficial outcomes may result from the submission of consumption data to safety assessors directly from individuals. Moreover, beneficial effects in terms of trust and confidence may emerge from a closer engagement of consumers in the risk assessment processes.

4.3 Blockchain and Food Traceability

Blockchain is deemed to have a disruptive impact in many sectors, including financial services, human resources, and intellectual property. The potentiality of this technology in the food safety domain is currently under discussion both in academia and the tech industry. On the one hand, academic studies have modelled supply chain traceability

² <https://etendering.ted.europa.eu/cft/cft-display.html?cftId=2827>.

³ <http://news.mit.edu/2017/artificial-intelligence-suggests-recipes-based-on-food-photos-0720>.

systems based on Blockchain and RFID [19]; on the other hand, IBM and Walmart have implemented two Blockchain pilots for the traceability of mangos and pork alongside the whole supply chain⁴. This research found several benefits both for consumers and the food industry. First, the improvement of track capabilities for the industry can result in a faster response for food recalls; secondly, the increase in transparency of the supply chain discourages frauds and illegal activities; thirdly, Blockchain is able to reduce compliance costs.

5 Emerging Issues for Data Governance

While the previous sections have identified current and future Big Data applications in the domain of food safety, this part introduces the emerging issues originated by these technologies for which forward-thinking policy solutions should be adopted.

‘Data governance’ consists of “the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets” [20]. This paper adopts a holistic approach defining data governance as “legal, ethical, professional and behavioural norms of conduct, conventions and practices”⁵. This perspective is grounded on the coexistence of the heterogeneous entities involved (academia, food industries, food safety authorities, and general public) that bring different views of governance together, thus calling for a broad definition of this term.

5.1 Data Sovereignty and International Food Data Exchange

Different approaches to the collection, the storage, and the use of data may result in non-inclusive policies implemented by regulative authorities belonging to different jurisdictions. While these conceptual resistances could lead to unintended forms of ‘data protectionism’, gathering data from multiple sources is crucial to perform a valid risk assessment through gaps minimisation and bias reduction. The global challenge raised by international information sharing requires the definition of peculiar regulations that satisfy all the stakeholders and comply with the internal rules adopted by each competent authority. Therefore, international information sharing platforms should be modelled according to access rules that balance the ownership of data with the general interest to an efficient risk assessment.

Alongside these practical reasons in support of international data exchange, ethical justifications could lead to a paradigm shift in the understanding of data sovereignty. On the one hand, global efforts aimed to the safeguard of the human right to adequate food and sustainability should be grounded on shared findings which could be implemented by joint data analysis. On the other hand, bearing in mind the absence of

⁴ [https://www-01.ibm.com/events/wwe/grp/grp308.nsf/vLookupPDFs/6%20Using%20Blockchain%20for%20Food%20Safe%202/\\$file/6%20Using%20Blockchain%20for%20Food%20Safe%202.pdf](https://www-01.ibm.com/events/wwe/grp/grp308.nsf/vLookupPDFs/6%20Using%20Blockchain%20for%20Food%20Safe%202/$file/6%20Using%20Blockchain%20for%20Food%20Safe%202.pdf).

⁵ <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf>.

technological infrastructures and the generally low levels of expertise, the sharing of information with emergent nations is desirable to reduce the digital divide between developing and developed countries.

5.2 Information Sharing and Competition in the Food Industry

Alongside cross-sectorial intellectual property regulations, both EU law and contractual agreements between data suppliers and EFSA grant the ownership of data to the originator [21], thus preventing a co-ownership situation or the entering into a license agreement. Previous research has observed that ownership can be broken into two facets, i.e. data protection and confidentiality. On the one hand, data protection applies to data submitted to support an application (e.g. to place a novel food on the market). It aims to protect the competitive position of the originator by denying the use of the same data for a subsequent applicant. On the other hand, confidentiality preserves the commercial value of data by not granting access to data submitted by the originator [22]. Two main arguments can be brought in favour of a different balance between transparency and secrecy of data.

On the one hand, the presence of concurring legislations that grant multiple layers of protection to data ownership might be reflected in uncertainties in the legal framework, thus leading to a self-restricting approach to the routine use of Big Data applications. As regards WGS, for instance, EFSA clearly concluded that “the legal and official systems are not yet adapted to the large-scale application of WGS to support food safety policies (and) legal obstacles are to be expected and a careful balance must be struck between the desirable complete openness from a food safety point of view and the privacy and related concerns that necessitate confidentiality” [12].

On the other hand, as regards the protection of the commercial value of data, the urgency of avoiding unfair monopolies by creating competitive markets for data calls for further investigation. French and German competition authorities have already underlined some important lines of research and identified potential solutions [23]. Their impact in the food industry should be discussed in light of the public interest in sustainability, transparency, and safeguard of fundamental rights related to food.

5.3 Privacy and Security Issues in Consumption Data Donations

The donation of non-personal data is a common practice in the food safety domain. ‘Calls for data’ are regularly published by risk assessment authorities to perform evaluations concerning fields where additional information is needed. The need of forward-thinking governance solutions and the entry into force of the EU General Data Protection Regulation⁶ (GDPR) raise the urgency of regulating personal data donations [24]. As shown, consumption data can be used to predict overweight and diabetes rates,

⁶ Regulation 2016/679 of the European Parliament and of the European Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L 119.

thus being subsumed to ‘health data’, one of the special categories of data subject to a stricter regime as regards data processing conditions and obligations under the GDPR⁷. The privacy of individuals involved in food safety pertains to the broader area of academic research that deals with the use of personal data in the scientific research, whose relevance is highlighted by the abundance of the literature on this topic.

On the one hand, as shown in the previous sections, individual data donations can be implemented as regards consumption data. The possible definition of data gathered from IoT, smartphones and social media under the category of health data poses privacy-specific questions and security risks due to the use of personal devices that can reveal sensitive information related to the data subject. On the other hand, developing a regulatory framework to encourage data donations from private companies to public bodies involved in food safety is a further challenge. Breaching data protection rules may expose the donors to liability, thus leading to an overprotective attitude that might limit donations. At the same time, the philanthropic nature underlying data donations should not suggest a less protective approach to safeguard measures, especially in light of the sensitive nature of processed data.

6 Final Considerations

Current Big Data applications are reshaping the way in which food industries, academia, food safety authorities, and consumers interact, collaborate, and fulfil their duties. Furthermore, possible future implementations present immense opportunities for these entities as regards the active engagement of all these actors. As shown, the flow of information between the food industry, consumers and food safety authorities raises open questions towards data sovereignty, competition, privacy within data sharing processes. Providing effective solutions to these issues is crucial to the implementation of the new applications discussed above. Maximising the positive outcomes of the cutting-edge technologies already in use requires clear data governance policies. Ultimately, the role of Big Data as a powerful instrument to protect and safeguard the fundamental rights pertaining to food and sustainability should be enhanced by policymakers.

Further research may explore two areas: on the one hand, bearing in mind the need of future-proof solutions, monitoring the development of future Big Data applications is crucial to understand to which extent new technologies interfere with stakeholders’ interests. On the other hand, further work is needed to understand how technical measures could enable all the involved entities to protect their interests without undermining the other stakeholders.

⁷ Recital 15 of the GDPR defines health data as the one “pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject [and] any information on, for example, a disease, disability, disease risk [...] of the data subject independent of its source”.

References

1. FAO: FAO and the SDGs Indicators: Measuring up to the 2030 Agenda for Sustainable Development (2017)
2. FAO: Voluntary Guidelines to support the progressive realization of the right to adequate food in the context of national food security, adopted by the 127th Session of the FAO Council (2004)
3. Strawn, K.L., et al.: Big Data in food safety and quality. *Food Technol.* **69**, 42–49 (2015)
4. Memorandum of Understanding Between the European Chemicals Agency (ECHA) and the European Food Safety Authority (EFSA) (2017)
5. Open Data Institute: Helping organisations navigate ethical concerns in their data practices (2017)
6. Beni, L.H., Villeneuve, S., LeBlanc, D.I., Delaquis, P.: A GIS-based approach in support of an assessment of food safety risks. *Trans. GIS* **15**(s1), 95–108 (2011)
7. Öcal, M., Kaya, I.A.: Food safety and GIS applications. In: 2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), pp. 85–90 (2015)
8. Cooley, M., et al.: Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PLoS ONE* **2**, e1159 (2007)
9. Decorte, L., Joris, I., Van Looy, S., Peelaerts, W., Bronders, J.: Software tool for calculating the predicted environmental concentrations (PEC) of plant protection products (PPP) in soil: final report. EFSA Support. Publ. **11** (2014)
10. FAO–WHO: Applications of Whole Genome Sequencing in food safety management (2016)
11. FSA: Chief Scientific Adviser’s Science Report Issue Three: Whole-genome sequencing of Foodborne Pathogens (2016)
12. EFSA Scientific Colloquium N°20: Whole Genome Sequencing of food-borne pathogens for public health protection (2014)
13. Kate, K., Chaudhari, S., Prapanca, A., Kalagnanam, J.: FoodSIS. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2014, pp. 1709–1718. ACM, New York (2014)
14. Lucas Luijckx, N., van de Brug, F., Leeman, W., van der Vossen, J., Cnossen, H.: Testing a text mining tool for emerging risk identification. EFSA Support. Publ. **13** (2016)
15. van den Puttelaar, J., Verain, M.C., Onwezen, M.C.: The potential of enriching food consumption data by use of consumer generated data: a case from RICHFIELDS. In: Proceedings of the 10th International Conference on Methods and Techniques in Behavioral Research, pp. 82–87. Dublin City University, Dublin (2016)
16. Shan, L., et al.: Interactive communication with the public: qualitative exploration of the use of social media by food and health organizations. *J. Nutr. Educ. Behav.* **47**, 104–108 (2015)
17. Regan, Á., Raats, M., Shan, L., Wall, P., McConnon, Á.: Risk communication and social media during food safety crises: a study of stakeholders’ opinions in Ireland. *J. Risk Res.* **19**, 119–133 (2014)
18. Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., Bell, D.: Analyzing the language of food on social media. In: Proceedings of the 2014 IEEE International Conference on Big Data, pp. 778–783. IEEE Press, New York (2014)
19. Feng, T.: An agri-food supply chain traceability system for China based on RFID & blockchain technology. In: Proceedings of the 13th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–6. IEEE Press, New York (2016)
20. Ladley, J.: Data Governance. Morgan Kaufmann, Waltham (2012)
21. Kocharov, A.: Data ownership and access rights in the European Food Safety Authority. *Eur. Food Feed Law Rev.* **5**, 335–346 (2009)

22. Holle, M.: The protection of proprietary data in novel foods. *Eur. Food Feed Law Rev.* **5**, 280–284 (2014)
23. Bundeskartellamt: *Autorité de la concurrence: Competition Law and Data* (2016)
24. Taddeo, M.: Data philanthropy and individual rights. *Mind. Mach.* **27**, 1–5 (2017)