The Alan Turing Institute

Roadmapping Uses of Advanced Analytics in the UK Food and Drink Sector

Final Report to the Food Standards Agency (FSA)

June 2020

Project FS301085

Rob Procter, Shuaib Choudhry & Jim Smith, Alan Turing Institute and University of Warwick

Martine Barons, University of Warwick

Adam Edwards, Cardiff University

Table of Contents

Executive Summary

This report is a state-of-the-art summary of the latest developments and issues in the application of advanced data analytics to food safety and authenticity, for improved system and consumer trust. The focus has been on identifying opportunities, barriers and enablers for the adoption of advanced data analytics by the food sector from the perspective both of private Food Business Operator (FBO) needs, and public-sector use and governance for the FSA as a food system regulator

The investigation has combined a literature review and 'live' exchanges with experts in data analytics and food sector practitioners, conducted via a series of interviews and a policy Delphi.

The report provides an overview of advanced data analytics methods, followed by a review of some applications of these methods reported in the recent research literature that were assessed to be of relevance to the FSA's mission. While the literature on applications of advanced data analytics in food safety is extensive, not surprisingly, the evidence of their translation into practice is harder to find. In this respect, the progress already made by the FSA suggests that it is well-placed to achieve its goal of transforming its capacity as a food system regulator and be able to meet the challenges that lie ahead.

We consider a number of issues that will need to be addressed if advanced data analytics are to be utilised in support of consumer interests and trust in UK food safety and authenticity. These issues are not specific to the FSA or, indeed, to the food sector in general. Instead, they are typical of those that any organisation/sector that is intent on using data analytics to innovate its processes must face. For some of these issues, such as explainability (which represents a major barrier at the moment), technical solutions can be expected to emerge within the next 2-5 years. Others, such as being able to share commercially sensitive data while preserving privacy demand, solutions and progress will depend on the capacity of stakeholders to reach agreement on principles of data governance, as much as on technical advances.

The report concludes with a series of recommendations for taking forward the application of advanced data analytics within the FSA in the light of currently understood food safety and authenticity risks and the study team's own understanding of FSA applications during the time of this review. These recommendations are then presented using a framework that is designed to help assess their readiness for developing and transitioning to operational deployment within the next 2-5 years. Data analytics methods are advancing at a very rapid pace, which makes it very difficult to assess their impact beyond the 2-5 year time frame of this review.

Glossary

Artificial Neural Networks: a biologically inspired computational architecture that learns from training data how to perform a task. There are various types such as convolutional neural networks.

Autoregressive Integrated Moving Average (ARIMA): a model fitted to time series data to better understand the data and aid forecasting.

Bayesian Decision Theory: a decision framework, informed by Bayesian probability, that aims to minimise some total expected risk that is pre-defined.

Bayesian Networks: a type of probabilistic graphical model that represents a set of variables and their conditional dependencies through the ‘nodes’ and ‘edges’ of a network.

Chain Event Graphs: an extension of Bayesian Network techniques where ‘nodes’ represent events and event trees are the underlying graphical model.

Cluster Analysis: grouping objects according to a common shared attribute(s).

Convolutional Neural Networks (CNNs): a type of deep neural networks primarily used in image recognition.

Error, Trend, Seasonality (ETS) Models: a time series modelling technique utilising exponential smoothing.

Food Business Operator (FBO): The natural or legal persons responsible for ensuring that the requirements of food law are met within the food business under their control.

Gaussian Processes: a probability distribution over possible functions and is an algorithm that can be used to carry out regression or classification when estimating an unknown function.

Geary’s C: a correlation coefficient that measures the overall spatial autocorrelation in data. Autocorrelation measures how one object is similar to others surrounding it and so is used in spatial analysis methods and geographic information science

Graph-based Methods: a family of tools for analysing the behaviour of systems that have a natural network representation. This includes looking for recurring patterns or anomalies.

Kriging: a Gaussian process regression method that is particularly used in spatial analysis.

Monte Carlo Methods: a broad class of computational algorithms that rely on random sampling to obtain numerical results and allow for the modelling of complex situations when there are many random variables involved and assessing the impact of risk.

Moran’s I: a correlation coefficient that measures the overall spatial autocorrelation in data. Autocorrelation measures how one object is similar to others surrounding it and so is used in spatial analysis methods and geographic information science.

Natural Language Processing (NLP): the application of computational techniques to the analysis and synthesis of natural language and speech.

Nowcasting: a forecasting method that aims to predicts variables in time ahead of the delay in the release of official data and statistics e.g. macroeconomic data or trade data.

Random Forest: an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decisions trees and outputting the mode of the decision trees or the mean of a regression prediction.

Seasonal Autoregressive Integrated Moving Average (SARIMA): a model fitted to time series data that accounts for seasonal trends, to better understand the data and aid forecasting.

Self-Organising Map: a type of artificial neural network that is trained using unsupervised learning to produce a representation of the training data.

Support Vector Machines: a supervised learning model that is used for classification and regression analysis.

Time Series: a series of data points indexed in time order. Often used for pattern recognition in temporal data as well as for forecasting and analysis to extract meaningful statistics.

User Generated Content: content created by people/users rather than a brand e.g. social media posts.

Aims and Objectives of the Investigation

The aim of the investigation was to provide the Food Standards Agency (FSA) with a state-of-the-art summary of the latest developments and issues in the application of advanced data analytics to food safety and authenticity. It is expected that the findings will be used to support the FSA in achieving its goals, i.e., “strategic prioritisation and translation of risks and opportunities” as it seeks to understand what steps “a competent central regulator” should be taking to advance its capabilities.

To achieve this aim, the report covers analytical method(s) that have reached a sufficiently mature stage of development that are likely to be available within the next 2-5 years to assist the FSA achieving its goal of delivering a safe and authentic UK food system. The objective is to better understand applications of analytical methods that are feasible now, and more challenging applications that are may be transformative in the near future and therefore worth considering further R&D.

Methodology

The project methodology consisted of three inter-related activities that were conducted in a six-month period between October 2019-March 2020[1].

Literature review: This was performed as part of the preparation for the subsequent interviews and policy Delphi. The aim was to (a) familiarise the project team, as technical experts, with the context of the FSA’s work, including current understanding of food safety and authenticity risks, practices for monitoring them, and to identify key informants for the subsequent stages; (b) summarise emerging data methods and match the analytical/technological opportunities they represent against known and foreseeable food sector pressures, including examples of applications from other sectors that might be transferable to the food sector.

Semi-structured interviews: Issues identified by the literature review were then followed up in a series of interviews with key informants. An initial list of interviewees was drawn up following consultations with FSA staff. This was supplemented by people identified through the literature review and suggestions from interviewees themselves. The interviews covered the following issues: 1) food safety and authenticity challenges; 2) applications of advanced data analytics to address these challenges. A total of thirty two people were interviewed: nine members of (a) the FSA and public sector organisations with interests in food safety; (b) twelve academics working on food safety research and/or in advanced data analytics; (c) eleven representatives of FBOs and professionals working for other bodies involved in food safety. Interviews were recorded and transcribed for subsequent analysis.

Policy Delphi: Interviews were followed by a policy Delphi (Appendix B), whose remit was to establish a consensus on recommendations for themes identified following analysis of the interviews. The policy Delphi is a well-established method for targeted

---

[1] Consultation with the FSA Data Team took place between October and November 2019. Several FSA Proof of Concept and Sprint projects have advanced since that time, in part resultant from COVID-19.

consultations with experts to identify consensuses for recommendations. Panellists take part in iterative rounds of deliberation, beginning with an initial questionnaire. Responses are synthesised and returned to panellists for a final round of evaluation. The method is designed to elicit the breadth of views that exist amongst informed respondents about a subject and consequently to maximise foresight about developments in this sector.

Prospective panelists from a long list drawn up following interviews were contacted and 19 (56%) responded, generating a panel that falls within the 10 – 50 range of respondents acknowledged by methodologists as sufficient to run a meaningful deliberation [1]. Of the 19:

- 1 was recruited from a Non-Governmental Organisation concerned with facilitating collaboration between government, business and academia in addressing issues of food safety;
- 2 were recruited from digital commercial organisations concerned with promoting the uptake of digital technologies by government and business;
- 2 were recruited from the government's food safety regulator;
- 7 were recruited from food business organisations concerned with the safety of food produce and its retail; and,
- 7 were recruited from academia and involved in researching trends in food safety.

Questions were defined for round 1 of the policy Delphi from the literature review and interviews and invitations sent to participants. Based on these responses, a second round of questions were distributed, and 14 responses were received. A summary of the findings can be found in Appendix B.

Ethical approval was received from the University of Warwick, Biomedical and Scientific Research Ethics Committee.

Summary of Food Safety and Authenticity Risks

Food safety and authenticity risks within the global food system are continually being shaped by a range of factors. The way the food system adapts will impact on known risks and potentially give rise to new risks or new sources of risk. Evidence from the literature, interviews and the policy Delphi were used to identify major food safety and authenticity challenges that the UK food sector is likely to face over the next 2-5 years. Below, we summarise these under four overarching and interrelated themes:

- **Environmental pressures;**
- **Changing consumer trends;**
- **Complexity of food supply chains; and,**
- **Sector governance and regulation.**

These themes reflect a range of underlying factors that are already evident or where the consensus is that they are likely to impact our food system within this period. Underlying factors of significance include:

- The global food system is exposed to “shocks” whose effect will be to undermine the resilience of the food supply chain, especially where these rely on just-in-time production and delivery [2]. These shocks range from growing geopolitical instability impacting on, e.g., shipping/trade routes, to pandemics with potential to impact across the whole food supply chain, as in the case of COVID-19[2].
- Regulatory divergence and increased pressure on the delivery of the UK food safety regulatory system. One important example is regulatory systems not keeping pace with new business models, e.g. ‘dark kitchens’ and ‘cloud kitchens’[3], online retail of food stuffs [3].
- Strain amongst food producers and traders facing multi-faceted pressures could lead to an increase in compromising practices. Also, increasingly, organised crime is infiltrating the food sector as a way of legitimising other illicit activities. This leads to for example, counterfeiting and tampering of food produce, economic adulteration and illegal labour practises [4].
- The growing global population and demand for food stuffs in increasingly competitive markets raises prices. This can stimulate economic adulteration, along with other fraudulent practices as producers and traders seek to cut costs [5].
- Climate-driven changes in food supply chains may force break-up of trading relationships and increase reliance on more geographically diverse (and possibly less transparent) suppliers [6].
- Changing consumer demands leading to product changes, including mass customisation. This creates pressures in production environments and increases risks of cross contamination, which is also a greater risk for food hygiene and allergen control, when food sensitivity and allergic reactions [7] are increasing in the UK population.
- Reducing our environmental footprint. New risks may accompany consumer demands to minimise plastic usage in packaging and changing consumption patterns, ranging from switch to plant-based diets, artisanal produce and novel products e.g. lab-grown proteins, insects, etc. [8, 9].

[2] The COVID-19 emergency broke during the second round of the Policy Delphi and, consequently, several references to its early potential impact were made in responses to the second-round questionnaire (see the final report of the Policy Delphi in Appendix B). However, this project has not reviewed the impact of COVID-19 on food sector data systems and how it may have advanced the application of data analytics, given the need for responsive ‘situational awareness’.
[3] Kitchens that do not have a traditional consumer facing service and produce meals exclusively for delivery.

Overview of Data Analytics Methods

Advanced data analytics covers a multitude of approaches commonly, if not always accurately referred to as Artificial Intelligence (AI). Data analytics paradigms can be classified into three basic types: supervised learning, unsupervised learning[4] and reinforcement learning. Supervised learning requires a training dataset of known examples of the phenomenon of interest (i.e. a labelled dataset). Examples of supervised learning methods include Bayesian methods, Decision Trees and Random Forests, Artificial Neural Networks (ANNs). As its name suggests, unsupervised learning does not require a training dataset and so can discover unknown patterns in data. Some examples are clustering methods, Self-Organising Maps Spectral methods and Spatial Correlation. Like unsupervised learning, reinforcement learning also does not rely on training datasets. Instead, the system ('agent') learns by interacting with its environment, optimising its behaviour from the responses it generates.

Bayesian Methods are well-established in industry and academia for inferring risk and modelling decision making processes. Good statistical methodologies have their Bayesian analogues and code is available via software libraries such as R.

Bayesian Decision Theory provides a formalism for decision making under uncertainty. Bayesian statistics are used to estimate the expected value of different actions and update expectations based on new information. The method can be used to aid decision making processes when trying to identify best courses of action during an incident or to pre-emptively prepare for incidents [10].

Bayesian Networks capture causal relationships in a probabilistic, graphical framework. They provide rigorous quantification of risks and decision modelling, and clear communication of results. Bayesian Networks have been utilised for risk assessments in health [11], manufacturing processes [12] and for issuing early warnings of natural hazards [13].

Chain Event Graphs are a recent and more powerful development of Bayesian Networks, allowing modelling of a chain of circumstances leading to an event and modelling of highly asymmetric decisions or processes [14]. Chain Event Graphs are already a viable tool for small scale problems, software is available through CRAN[5] and applications are now beginning to appear. However, finding ways of scaling up model selection procedures so that Chain Event Graphs can provide real time analysis for problems with more than 15 variables is a major challenge. New tools are currently being developed to manage this problem, but they are unlikely to be ready for operational deployment within the next 2-5 years.

All Bayesian methods need a prior, which quantifies the belief in the probability of some event in advance of new information being available. This can be elicited from domain experts. If this is not possible, computational techniques, such as Gaussian

---

4 Semi-supervised learning, which falls in between supervised and unsupervised learning, combines a small amount of labelled data with a large amount of unlabelled data during the training phase.
5 https://cran.r-project.org/web/packages/ceg/index.html

process methods, can be used instead. As new information becomes available, priors can be updated and then used to refine the parameters of the underlying mathematical model of the process.

Explainability refers to the capacity of a machine learning model to make its behaviour transparent or understandable to its users. 'White box' methods such as decision trees and Bayesian methods have good explainability, whereas 'black box' methods such as ANNs and deep learning techniques generally do not. Explainability is increasingly recognised as a key requirement if data analytics tools are to be trusted by their users as aids in decision-making tasks [77].

Box 1: Explainability and machine learning

Decision trees are a very well-established and widely used supervised learning method of model building [15]. One of their advantages is that it is easy to extract rules that explain the behaviour of the model. Random forests are a development of decision trees that improve their accuracy [16].

Graph-Based Methods are a family of powerful tools for analysing the behaviour of systems that have a natural network representation. They can be used, for example, to detect anomalous patterns in networks that may be signals of unusual or suspicious behaviour. Spectral methods are emerging as particularly effective approaches that are also computationally efficient [17]. Network techniques have also been used to simulate the likelihood and spread of diseases in livestock and how best to control/intervene, through the routine collection of livestock movement data [18]. Social network analysis techniques have been used to model the transmission pathways of highly communicable diseases such as norovirus [19].

Artificial Neural Networks (ANNs) are another family of machine learning methods with a diverse range of applications. Deep learning methods (e.g., convolutional and recurrent ANNs), which are variants of multi-layer ANNs, are finding application within all three machine learning paradigms and offer significant performance improvements in applications such as machine translation [20], image classification [21] and speech recognition [22]. Self-Organising Maps are another type of ANN that has proven useful for dimensionality reduction, where non-significant variables are eliminated.

Natural Language Processing (NLP) covers a wide range of techniques for the extraction of information from unstructured data. Applications include opinion mining [23], event detection [24] and rumour verification [25]. NLP is one area of advanced data analytics where deep learning (i.e., convolutional and recurrent ANNs, including, Long-Short Term Memory (LSTM)) is making a significant impact [26]. Recent advances also include using word embeddings to improve information extraction by capturing the semantic similarities between words [27].

Spatial Processes, where patterns are studied within a spatial context, are an extremely well-studied and utilised model in areas such as public health [28]. Spatial patterns and correlations play a vital role in epidemiological processes [29] and in human geography [30]. Examples of methods for finding patterns and correlations in spatial data include Moran's I or Geary's C [31]. Once spatial patterns have been identified, more advanced analytics can be performed, such as Kriging methods, which utilise gaussian processes to aid prediction by taking into account the spatial structure of the data and then perform spatial interpolation [32]. Using travel data combined with population demographics and pathogen information one can predict the risk and introduction site of exotic plant, human and animal pathogens. A risk ranking can then be used to develop pathogen surveillance programs [33].

Monitoring of Food Safety and Authenticity Risks and the Application of Advanced Data Analytics

The FSA collaborates with enforcement partners in Local Authorities and Port Health Authorities to acquire sampling and compliance data from inspections and testing of food commodities as they enter and travel through the UK food supply chain. Local Authorities also conduct inspections of restaurants, takeaways and food retailers for compliance with hygiene directives. However, with Local Authority resources under pressure, this regulatory delivery system is under significant strain, which emphasises the importance of being able to prioritise resources. Inspections and testing are costly, so the effectiveness of the system depends on inspections being targeted where there is likely to be the greatest risk. Testing also takes time, which may limit the effectiveness of any mitigation measures such as recalls, particularly for the last-minute supply chains for many perishable goods. A recent NAO report found that spending by local authorities on food inspections has fallen in the past decade by 20%. While the number of food hygiene inspections were broadly stable, authenticity checks fell significantly during this period [34].

Two conclusions follow from this: first, inspections need to be driven by intelligence that enables the accurate identification and analysis of risk, where possible before it has entered the UK food system i.e. includes predictive capability; second, intelligence capacity needs to be available at key intervention points throughout the food supply chain.

The FSA Strategic Surveillance team has trialled the use of a range of new data analytics techniques for the identification of risk through its ongoing programme of 'Proof of Concept' (PoC) and 'Sprint' projects (see Appendix A).

In this section, we review where advanced data analytics are likely to have the greatest impact towards the FSA achieving its objectives.

The use of social media as an 'observatory' is a well-known surveillance technique, which has been applied in the early detection of a wide range of incidents and events [35], as two FSA PoC projects have illustrated. However, the noisy character of social media makes information extraction difficult to achieve with an acceptable level of accuracy. Recent advances in NLP-based information extraction techniques are beginning to deliver more robust performance [36-37].

Food processing plants, such as abattoirs provide a key intervention point for many food safety and authenticity risks. Sensor technologies in abattoirs, may offer the opportunity to augment safety inspections but also improve for example animal welfare standards. Recent work has utilised deep learning methods to analyse audio signals and one application has been in bioacoustic monitoring of animal welfare in farms [38] and abattoirs [39]. Convolutional ANNs have been applied to image data to recognise animal behaviour from video data [40-41] and to determine the quality and safety of meat products [42].

Food fraud (intentionally debasing the quality of food offered for sale either the admixture or substitution of inferior substances or by removal of some valuable ingredient) typically exploits vulnerabilities earlier in the food supply chain. Several different techniques have potential here, and some of these feature in the FSA's data science team's own work in this area.

The ISAR (Import Screening for the Anticipation of Food Risks) tool, developed for the Bavarian State Office of Health and Food Safety, uses import/export data from 2400 different food items to build seasonal, time series analysis models for inferring food commodity fraud risk [43]. On the available evidence, it appears to have been validated on just two cases of food fraud (melon seeds and hazelnuts). FSA data science team engaged with the Bavarian State Office of Health and Food Safety in the development of ISAR, and this translates to its own work on Risk Likelihood and Signal Prioritisation Dashboards.

Random forests have been applied to predict supply chain disruptions, which can then be used to infer the likelihood of criminal activity [44]. Monte Carlo simulations have been used to explore different scenarios in order to determine best course of actions and see possible effects of different mitigation strategies [45]. Reinforcement Learning and Monte Carlo control have also been utilised to develop context dependent response policies for disease outbreaks.

Graph-based methods have demonstrated their value for the study of the behaviour of food trade networks [46]. They have also been used to study the impact of shocks, such as COVID-19, on food trade networks and the associated policy implications [47]. Recent work has applied graph-based methods to the identification of food commodities at risk of fraudulent behaviour[6]. Analysis of network structure can provide information on how food safety risks may propagate across the supply chain and can be used, for example, to determine the source of a foodborne disease outbreak [48].

Image data, including satellite imaging is widely used to monitor the condition of food commodities at source. Self-organising maps have been used to predict fungal infections in herbs [49] and disease in wheat [50]. Support vector machines (a type of ANN) have been used to aid image recognition of parasites and pests in strawberry greenhouses [51] and seed borne disease in rice [52]. Deep learning methods such as convolutional ANNs have been used for plant disease detection and diagnosis [53]. ANNs have been used with satellite imagery and soil data to predict yield of crops [54] and large food retailers are now deploying such techniques on a range of commodities[7].

Large food retailers are using data harvested from the Web to pick up trends that might indicate increased risks, such as triggers for a poor crop yield that might incentivise adulteration or falsification in/of the country of origin. Other applications of advanced data analytics include monitoring certification compliance. For example, Oceanmind

6 Interview with Data scientist, November 2019.
7 Interview with large food retailer.

uses satellite imagery and vessel monitoring systems to check whether fish catches comply with MSC certification, something that was previously impossible [55].

The falling costs of whole-genome sequencing (WGS) technologies is leading to its increasing use by large food retailers and government agencies [56]. Applications include investigation of foodborne outbreaks and surveillance to delineate local, regional and global genomic epidemiology of pathogens and to attribute the infection source. For example, WGS has been used by Public Health England to trace a Salmonella outbreak to a single egg producer [57]. Machine learning is now being used to accelerate the WGS analysis process. For example, Trace Genomics[8] uses machine learning applied to the soil microbiome to test for pathogens, yield productivity and overall soil health. Bayesian methods have been employed to determine disease likelihood in crops and animals [58]. Machine learning on WGS data also has applications in food manufacturing, where indigenous bacteria can hamper pathogen detection and limit shelf life [59].

Tracing the source of foodborne disease outbreaks and determining their likely impact are key for devising mitigation strategies [60]. A Random Forest classifier has been used to model the different molecular structures that characterise the different Salmonella strains with differing strengths, impacts and severities [61]. Gradient boosted tree models have been applied to genomic data to predict the antimicrobial resistance of pathogens [62].

Deriving predictive knowledge from genome data is still a nascent field and it has primarily been used for taxonomic and comparative purposes. It should also be stressed that access to well-curated genomic databases is essential in all these applications [63].

'Horizon scanning' is an important tool for early warning of unanticipated or hitherto unknown risks. The European Food Safety Authority (EFSA) has developed the Emerging Risks Identification Support System (ERIS), which utilises information extraction from a range of sources, including the scientific literature, to support the early detection of emerging safety risks in supply chains [64]. To the best of our knowledge, to-date, ERIS has been tested on just two fish farming supply chains.

The United Nations Food and Agriculture Organisation (FAO) EMPRES Global Animal Disease Information System (EMPRES-I) also aggregates and extracts information from a range of sources, including national and regional reports and databases, to provide search-based services on animal diseases [65].

8 https://tracegenomics.com/

Challenges for the Adoption of Advanced Data Analytics

Analysis of the information gathered reveals a number of distinct challenges that the FSA will need to address if it is to progress its use of data analytics in food safety and authenticity risk monitoring.

Datasets

The key to the success of data analytics is the availability of datasets that are appropriate to the task at hand, are of good 'quality' and are accessible in a timely way.

Data quality covers a diverse set of issues, including missing data, anomalous values, source integrity, non-standard formats and non-stationarity (i.e. changes in data properties). These issues are not specific to data analytics but may become more challenging to deal with in this context. This is particularly the case for so-called 'big data', where factors such as volume, heterogeneity and velocity (i.e. speed of arrival) may mean that established procedures for ensuring data quality may not scale well. We pick up some of these issues below, but a comprehensive review is beyond the scope of this report.

The FSA commonly uses open data sources, such as administrative records from trade databases or climatic data. Administrative data has proven of value in a number of applications (e.g. analysing trade patterns) but can be less useful for rapid-response risk surveillance, if potentially by the time data is available, fast-moving supply chains may mean consumers have been exposed to the identified risk. Data on financial transactions is emerging as a potentially valuable source for detecting food crime using techniques such as anomaly detection but getting FBOs to agree access is problematic. Government departments such as HMRC also hold potentially valuable data, but progress on inter-departmental data sharing can be similarly complex.

Problems in finding and getting access to datasets in a timely way came up frequently in interviews and in Delphi panellists' responses. The meaning of 'timely' is, of course, dependent on the use being made of the data and the time window for effective mitigation should a problem be detected. In general, that time window is narrow in the food supply chain. One example where the timeliness requirement is not currently satisfied is the monitoring of trade flows: the data held by Comtrade[9], which is a UN International Trade Statistics Database, has a lag of 3 months and sometimes much longer, making it unsuitable for the early prediction of food fraud or where there is an abrupt change of conditions, such as in the COVID-19 pandemic. One way to circumvent this kind of timeliness problem is to use nowcasting[10] methods as a way of providing an estimate of the data before its release (see Box 2) [66-7].

While there is no doubt that more and more valuable data is being generated as a by-product of the food supply chain, the problem that remains for the FSA is securing access to it. Though not fundamentally addressed by it, still requiring permission agreements, it is possible data access may be improved by the advent of 'digital twins'.

---

9 https://comtrade.un.org/
10 i.e. predicting a current value rather than a future one.

The aim of which is to use ‘Internet of Things’[11] (IoT) technologies to create a digital replica of the food supply chain and may help remove the ‘burden’ of some information exchange. While this is technically possible as a proof of concept, there are many challenges to be overcome before this is likely to be realised in practice: as Delphi panellist observed, the food supply chain is fast moving, rapidly changing and extremely complex.

> Many official datasets have significant delays in the publishing of new data, which limits their value in a number of application areas. The use of nowcasting techniques to provide estimates of current values, offers a potential solution to this problem and have proved their value in, for example, public health planning and control applications [68].
>
> User Generated Content (UGC), such as social media and internet search logs, are examples of new data sources that have nowcasting potential as they are high-frequency, real-time and may be reasonably easy to harvest. However, some UGC is also very noisy, making it challenging to extract a good signal.
>
> It should be noted that, as always in nowcasting, there is a trade-off between timeliness and robustness.

Box 2: Nowcasting

Among these challenges are standards to ensure data quality and interchange (see below), interoperability of devices and mechanisms to reduce risks of data tampering and fraud. For the FSA, access to data from within the food supply chain would be transformative for the early detection of a range of food safety and authenticity risks.

Standardised procedures and formats for collecting and recording data will be essential if data from different sources is to be aggregated and linked easily, and the effort expended on data preparation minimised [69]. However, these are often not yet in place and will take time to agree. It is also important to bear in mind that, while it may be technically possible to guarantee data integrity once input, verifying integrity at source remains a problem with no obvious technical solution. For example, if the sources are sensors, human oversight will be needed to verify they are being deployed appropriately.

Because of its commercial sensitivity, FBOs have often been reluctant to share much of their data. However, there are some modest signs of progress. For example, in the past six months, the FSA’s National Food Crime Unit (NFCU) and FSS (Food Standards Scotland) have gained access to sampling data provided through the Food Industry Intelligence Network (fiin)[12], which is a body set up by major UK FBOs, including retailers, manufacturers and service companies to share intelligence on food

---

11 Networked devices capable of sensing, sending and receiving data.12 https://www.campdenbri.co.uk/news/fiin.php

12 https://www.campdenbri.co.uk/news/fiin.php

authenticity and traceability. fiin is currently developing web-based, a cloud-based platform for sharing data between its members.

UGC is another potentially valuable new source of data. The COVID-19 pandemic has exposed public concern about the collection of personal data, even where there may be a strong policy case for doing so, reinforcing the importance of openness with the public where data sources may contain identifiable personal information.

Data preparation absorbs a lot of time and resources in any data analytics project and this can be further exacerbated by any lack of agreed data standards, which is a common problem with trade data for example. 'Data wrangling', i.e., dataset cleaning, preparation, organising, integrating, etc., has been estimated to take up to 60% of the time to deliver a data analytics project [70].

Some new data sources may pose significant quality challenges. UGC, for example, is often very 'noisy'. Problems may arise even when dealing with well-curated datasets that are of good quality, such as that from the International Food Safety Authorities Network (INFOSAN) and other administrative sources. In addition, some Delphi panellists expressed doubts about the quality of data that may be sourced from FBOs unless they are incentivised or legally obliged to do so.

When developing new data analytics applications, datasets of known value (e.g., trade flows, etc.) will need to be assessed for their quality, granularity and timeliness and the value of new sources of data evaluated. Feature engineering, the process through which attributes of datasets relevant to a specific application are identified, remains a largely manual exercise. Supervised data analytics methods require training data, which may be difficult to obtain and costly to generate.

The Artificial Intelligence for Data Analytics (AIDA) project at The Alan Turing Institute (ATI) is starting to develop some practical ways to reduce this effort but this and other projects are still in the early stages of delivering production-ready solutions [71]. The AIDA project has also developed a framework for thinking through the main issues of data wrangling.

Technical Infrastructure

More advanced data analytics methods and potentially huge volumes of data will make greater demands on compute and data infrastructure. In general, compute requirements are greatest at the model building stage but are much lower when models are subsequently deployed. For this reason, on-demand, scalable services offered by cloud computing have become the preferred option for many data analytics applications [72].

Distributed Ledger Technology (DLT) (often known as Blockchain, though this is just one version of DLT) has been proposed as the solution to the needs of FBOs, regulators and consumers for traceability within the food supply chain [73]. A key principle of DLTs is that a record should be immutable once added to the ledger, thus guaranteeing record integrity. However, the computational requirements for guaranteeing immutability in so-called "trustless" DLTs limit the rate at which records can be added. While this may not be a problem for low volume transactions, it makes them impractical at present for applications requiring a high record transaction rate [74]. Adopting a "layered" approach to DLT architecture may provide a solution to this problem [75]. However, it should be noted that there may also be issues about data privacy in trustless DLTs, where participants may be in competition with one another, a point emphasised by policy several Delphi panellists. One further note of caution relates to the proliferation of different architectures for DLTs, which is very likely to lead to interoperability problems in the future, with the risk that data becomes siloed [73].

Data trusts, i.e. “a legal structure that provides independent stewardship of data” [76] offer a socio-technical solution to these problems: participants agree who is allowed to add records to – and read records in – the ledger. Naturally, reaching agreement on which participant(s) are to be trusted in this way may not be straightforward. It should also be noted that relaxing this core principle makes the advantages of DLTs over a centralised, encrypted database less clear, with the advantage that data trusts can be agnostic about the choice of the underlying technology. The Open Data Institute (ODI) has undertaken a number of data trust pilot studies, including one with Food and Drink Manufacturers and the FSA is currently collaborating with the Internet of Food Things Network (IoFT) on an additional Proof-of-Value project. While Delphi panellists generally agreed on the technical feasibility of IoFT (subject to interoperability standards being agreed), several commented on potential barriers to adoption, including industry concerns about cost and surveillance.

Type 1 tools provide facilities (a “dashboard”) to visualise quantitative, time series data of known value for analysis. Deriving insights relies on the user’s expertise in understanding the significance of the patterns revealed by the dashboard. An example is the FSA PoC project “Understanding olive oil trade patterns”.

Type 2 tools visualise information extracted from unstructured data. An example is the FSA PoC project “Understanding allergy related discussions using social media”. These tools serve as an “observatory” by providing a count of mentions of known allergens. Another example is the FSA PoC project “Food/hazard extraction from media articles”, which uses information extraction from mainstream news media. The distinction between Type 1 and Type 2 tools stems from the latter’s information extraction step; the data generated is typically less reliable and thus requires greater effort in its interpretation.

Type 3 tools add a simple and well-understood predictive model to the basic dashboard capability. An example is the FSA PoC project “Predicting Vibrio infections from climate data”, which uses the known correlation between surface sea temperatures and Vibrio infection rates to highlight periods of increased risk.

Type 4 tools lack an explicit model of the underlying phenomenon and so use data analytics to learn (induce) one. For example: the FSA is currently conducting a PoC project to explore applying deep learning tools to model safety risks in restaurants and drive ‘intelligent’ inspections; the FSA has also completed a project to predict aflatoxin risks. Another example is learning models to detect “Unregistered food businesses”.

A list of FSA PoCs and sprint projects (as of November 2019) can be found in Appendix A.

Box 3: A taxonomy of data tools and their explainability characteristics

Explainability

The FSA data science team has found that users are don't like tools they perceive to be 'black boxes', i.e. whose behaviour is difficult to understand. This is consistent with views expressed by interviewees and in a recent report from ATI on ethical frameworks for AI [77], which states that explainability (Box 1) is a key requirement for the successful application of data analytics. More broadly, the ATI report argues that explainability is one part of a range of requirements that applications of data analytics should satisfy if their use is to be transparent and fair.

Given the current state of the art in explainable AI (xAI), this would limit the use of data analytics tools to those that have either: (a) been developed around known causal models (e.g., vibrio incidence); or (b) where a causal model may be derived either through (i) a priori knowledge elicitation process with experts or (ii) a post-hoc knowledge elicitation from models learned from data. This therefore places limits on the current value of some more advanced deep learning analytics. While how to incorporate explainability into this latter class of methods is a very active area of research, with contrastive and counterfactual techniques among the most promising lines of investigation [78-9], viable techniques are still some way from being identified.

Box 3 is a taxonomy of data analytics based on their inherent explainability capabilities. Based on this, the low explainability barrier of Type 1-3 tools make them good candidates for "low hanging fruit" applications. In contrast, explainability techniques are not yet available for Type 4 tools, making their deployment problematic, at least in the near term.

Skills

The rapid pace of the development and adoption of data analytics is already beginning to make its presence felt in the jobs market in the UK and globally, with shortages of skilled people widely reported [80], a point reinforced by many Delphi panellists. As it considers how to increase data analytics capability and capacity, the FSA, like other public sector organisations, will face stiff competition from private sector businesses.

Change Management

On the balance of the evidence available, the automation food safety and authenticity risk decision-making is an unlikely and undesirable outcome of the adoption of advanced data analytics. The users of these new analytics tools will need training and the opportunity to discover their strengths and weaknesses if they are going to be able to employ them effectively. Decision-making procedures are also likely to have to adapt if the FSA is to be able to manage and respond in a timely and effective manner to a growing volume of risk signals. We are aware that the FSA data science team is putting significant effort into change management as its projects move on from the PoC stage into deployment. Embedding new technologies within organisations takes time and so it will be important to continue tracking how new tools are being used and using what is learnt in a process of continuous improvement.

Readiness for Adoption

Any plans for the application of advanced data analytics for food safety and authenticity risk monitoring and mitigation by the FSA need to be considered in the light of a number of factors that may impact the prospects for success. Based on the evidence gathered in this study, a list of relevant factors is given below:

Strategic Value: address food safety or authenticity risks that are important to the FSA as a food system regulator and are: (a) currently a known gap; or (b) predicted to change within next 2-5 years; (c) have high impact as measured by (i) severity and/or (ii) scale;

Dataset Availability: (a) official, open, good quality; or (b) official or proprietary, negotiable access, good quality; (c) harvestable, requiring little cleaning;

Ethical and Legal Compliance: methods and development processes that can demonstrate compliance with the recently published ATI ethical framework, i.e., (a) process and (b) outcome transparency, including explainability;

Opportunity Score: a summary of factors 1-3 reflecting potential benefits, e.g., (a) bringing enhancements to existing practice, including improving robustness, enabling more timely interventions and/or reducing costs; or (b) establishing new practices and competencies;

Method Availability: maturity of methods as evidenced by demonstrated practical value in similar applications and supported by high quality software tools;

FSA PoC or Sprint Projects: well-defined use case, satisfactory results in terms of performance and so qualify as potential minimum viable product;

Generalisability: have potential to be applied to other use cases with minimal additional effort.

Analytics Score: a summary of factors 5-7, reflecting operational readiness of the method(s).

Overall: a summary of opportunity and analytics scores to present an overall indication of the readiness for adoption.

For each of these criteria we propose a three-point scale: 1 – not met; 2 – likely to be met within 2-5 years; 3 – met now. These are then used to score the each of the recommendations listed in Section 8 (see Table 1).

Recommendations

Below, we summarise a series of recommendations that we conclude will help facilitate the FSA in taking forward its goal of achieving a transformational impact on food safety and authenticity.

The recommendations are based on a synthesis of the evidence gathered from the literature review, interviews and two policy Delphi rounds.

Each recommendation begins with a statement in generic terms of the application or methods the recommendation endorses. This is then broken down into one of more specific projects that would enable the FSA to make progress in meeting it.

1. **Enhanced dashboards**: building on PoC and Sprint projects.
    a) Type 1 dashboards, such as “Understanding olive oil trade patterns”, and Type 3 dashboards, such as the “Predicting aflatoxin risks” dashboard, have the potential to be generalisable to other food commodities.
    b) Type 2 surveillance dashboards, which extract information from unstructured text, such as “Understanding allergy related discussions using social media”, “Detection of Norovirus outbreaks from Twitter posts” and “Food/hazard extraction from media articles“, would be enhanced through the use of advanced NLP-based information extraction methods to reduce false positives and negatives. Deep learning methods using word embeddings, for example, have advanced the state-of-the-art significantly [81]. While acknowledging the challenges of using high frequency, noisy data sources, advanced NLP information extraction methods may also have value for nowcasting of, for example, foodborne and infectious disease outbreaks [82].
    c) The same Type 2 dashboards could be promoted to Type 3 by adding prediction capabilities. For example, named entity recognition [83] and extraction of location information [84] would enable the use of spatial analytics to trace sources and to predict the spread.

2. **Horizon scanning**: early warning of unanticipated and unknown risks.
    a) Information extraction from scientific literature and reports on food safety can provide support horizon scanning for the early detection of new and emerging potential risks. More generally, this would support the conduct of systematic reviews for learning lessons from past food safety incidents [85].

       Such methods are already used by companies offering risk assessment services to businesses, including FBOs. This is the kind of application where collaboration with an external partner may be preferable to pursuing an in-house solution. For example, the National Centre for Text Mining (NaCTeM)[13] has developed a suite of tools that could be customised to deliver this kind of capability.

13 http://www.nactem.ac.uk/

3. **Intelligence-driven inspections**: better tools for a range of inspection scenarios.
   a) Intelligence-driven inspection was the focus of a recent PoC project, which used records covering multiple hygiene factors, physical conditions of the business and other variables with a convolutional ANN. There is scope for building on this by experimenting with different deep learning models (e.g., recurrent ANNs) and additional datasets. Regarding the latter, while noting quality challenges, Delphi panellists were agreed that there would be value in exploring the use of signals extracted from UGC sources.
   b) 65% of 'dark kitchens' are estimated to be unregistered (compared to 98% percent registration for traditional restaurants) and no effective system exists to verify their location. A recent PoC project used data from websites to build a classifier to assist in their detection. No details of the methods used or of the performance were available at the time of writing, but it is likely there would be scope for developing this further by experimenting with deep learning methods and new datasets.
   c) Application of deep learning for intelligence-driven abattoir inspections using acoustic and image data. There are many candidate methods. A survey can be found in [41, 86-7].

Any intelligence-driven inspection system must be seen to be fair and transparent by both FBOs and the public. Again, before any tools may be employed, explainability must be satisfied.

Bayesian methods provide a natural way of encoding causal relationships and so offer the means to satisfy the requirement for explainability. Many current machine learning methods have their Bayesian analogues. As other explainability techniques are developed, a wider range of data analytics methods, including deep learning, will also become viable options.

4. **Detection and Prediction of food crime**: generalisation and development of more advanced tools.
   a) FSA projects in this area have focused on a specific food commodity that is a known fraud risk. They have potential for generalisation in two distinct ways: (i) application to other known food fraud risks; (ii) for the discovery of fraud risks in commodities previously not suspected.
   b) These tools look for unusual trends or anomalous patterns in trade data that may be signatures of food crime. Time series models using trade flows and commodity prices have shown potential [43]. Earlier detection would benefit from use of more timely data, for which nowcasting offers a potential solution. An example is the use of ship tracking data to generate near real-time information on trading networks that underlie supply chains [88-9]. Bayesian methods have demonstrated good performance in a wide range of nowcasting applications.
   c) Advanced predictive capability for food crime requires powerful methods together with a wider range of datasets, e.g., INFOSAN, financial transactions, commodity prices, weather conditions, infectious disease outbreaks, supply

chain disruptions and geopolitical unrest. Bayesian Networks and Graph-based methods would be appropriate choices.

5. **Mitigation of food-related incidents**

A key element of mitigation response and planning is the identification of the source of incidents such as foodborne disease outbreaks among consumers [90-91], animals [92] and predicting spread and severity.

a) Spatial analytics techniques can assist in the identification of sources for mitigation response and in forecasting the likely incident spread for mitigation planning [92].

b) Bayesian networks are one of several network and Graph-based methods that can be used to predict how food safety risks may propagate across a network [93] and can be used to determine the source of a foodborne disease outbreak [94].

c) The falling costs of WGS is opening up opportunities for its use in food safety and authenticity risk management. However, WGS data is of high dimensionality[14], which makes analysis using conventional methods time consuming. Machine learning provides a solution to this problem and enables accounting for individual effects that are dependent on interactions with other genetic and environmental factors [95].

   A range of machine learning methods have shown promise for increasing the speed and accuracy of WGS data analysis for tracing outbreaks of foodborne diseases, including Support Vector Machines, Random Forests, Bayesian, ANNs and K-means clustering [96-7].

   Decision trees have been used with WGS data to predict pathogen antimicrobial resistance [62]. Bayesian methods have been used with WGS data to predict properties such as risk of diseases in crop and animals [93] and human health risk at the population level from foodborne pathogens.

6. **Modelling food system resilience**

Experience gained from projects outlined above provides the foundations for more ambitious projects. Assuming that access to good quality, timely datasets from across the food supply chain continues to improve, then it will be feasible over the next 5 years to model the behaviour of the food supply chain and its response to different kinds of ‘shock’. This would have significant value for predicting and mitigating the impact on food safety and authenticity, and for policymaking for improving supply chain resilience. However, this would require cross-government collaboration.

Access to a wide range of datasets, including food industry datasets, will be important. The COVID-19 pandemic may help to incentivise FBOs and government departments towards a greater sharing of data, as well as encouraging closer cooperation in developing tools.

[14] Commonly referred to as “the curse of dimensionality”.

Table 1 is a summary of how the criteria identified under the section on readiness for adoption may be mapped onto to these recommendations. The 'Overall' column grades each challenge in terms of its readiness to be taken forward within the next 2-5 years on the assumption that the criteria are equally weighted. We have used our own judgement, based on what we have learnt during the conduct of this study, in choosing the values assigned to criteria such as strategic value and opportunity score. However, we recognise that the FSA is much better placed to decide accurate scoring. We hope to provide a readily accessible means to articulate deployment opportunity to others.

It should be noted that, in many cases, it is not possible to recommend a specific data analytics method as there are a number that may be applicable. This means that experimentation will be necessary to identify the best performing.

Many statistical methods enable modelling (with explanations but a degree of speculation) of parts of a system where data is not retrievable or is costly to obtain. Hence, it may be important to position new data analytics methods, such as deep learning and natural language processing, so that they complement current methods.

Bayesian methodologies are one way to deliver this. Embedding expert judgment in terms of setting proper prior distributions can have very helpful effects in tightening the inferences that can be made from a particular dataset. By removing possibilities that are implausible for a given domain, data centric evidence can be much more informative and may help tighten confidence bands significantly. The key issue is the availability of such judgements and the need for transparent guidance where these exist.

Finally, as with data analytics applications generally, there will be a need to regularly re-calibrate – and sometimes rebuild – models as the behaviour of the food supply chain and the actors that constitute it change.

Potential applications for advanced analytics	Strategic Value	Dataset Availability	Ethical and Legal Compliance	**Opportunity Score**	Method Availability	PoC or Sprint	Generalisability	**Analytics Score**	**Overall**
Enhanced dashboards	3	3	a:3 b:2 c:2	a:3 b:2-3 c:2-3	a:3 b:2 c:2	a:3 N/A N/A	a:3 b:3 c:2	a:3 b:2 c:2	a:3 b:2 c:2
Horizon scanning	a:3	a:2-3	a:3	a:3	a:3	N/A	a:3	a:3	a:3
Intelligence-driven inspections	3	a:2-3 b:2-3 c:2	2	a:2-3 b:2-3 c:2	2	a:3 b:3 c:2	a:2 b:1 c:1	a:2-3 b:2 c:1-2	a:2-3 b:2 c:1-2
Detection and prediction of food crime	b:3	b:2	b:3	b:2-3	b:2-3	b:3	b:2	b:2-3	b:2-3
Mitigation of incidents	a:3 b:3 c:3	a:3 b:3 c:2	a:2-3 b:3 c:3	a:2-3 b:3 c:2-3	a:3 b:3 c:3	N/A	a:2 b:2 c:2	a:2-3 b:3 c:3	a:2-3 b:3 c:2-3
Modelling food system resilience	3	2	2-3	2-3	2-3	N/A	N/A	2-3	2-3

Table 1: Classification of project readiness (3 – met; 2 – met within 2-5 years; 1 – not met. Prefixes a, b, c, etc., refer to the labelling of projects within each recommendation)**.** Factors contributing to both the 'Opportunity Score' and 'Analytics Score' combine to reflect the overall readiness for adoption.

Note: Where a range of values is shown, this reflects how they may be expected to depend on which of several possible projects under each heading is selected.

Acknowledgements

We would like to thank the Food Standards Agency and its independent Science Council for supporting this study. We would particularly like to thank Ben Goodall for his invaluable guidance throughout this project. We would also like to express our sincere appreciation to all interviewees and policy Delphi panellists for their generous contribution of their time and expertise.

Appendix A: List of FSA Proof of Concept (PoC) and Sprint Projects

Understanding Olive Oil trade patterns and anomalies

Predicting Vibrio infections using climate data

RASFF analysis to identify UK specific hazards

Identifying anomalies in global trade patterns with the UK

Understanding allergy related discussions using social media

Online display of FHRS ratings

Meat mass balance analysis for detection of anomalies

Risk likelihood dashboard

Climate risk (aflatoxins)

Climate risk (pesticides)

Meat establishments dashboards

Detection of unregistered FBOs

Signal prioritisation dashboard

Using non-UK RASFFs to predict UK RAFFs

Trade routes and volumes

Predicting incidents of wine fraud in Europe

Consumer attitudes to food pricing extracted from social media

Appendix B: Summary of Policy Delphi Findings

Introduction

This appendix sets out the key findings of the policy Delphi deliberation undertaken as part of the research into the uses of advanced data analytics in the UK food and drinks sector. It provides a more detailed summary responses.

Overview

This policy Delphi drew upon initial interviews with key informants in the regulation of food standards, FBOs and in academic research to identify 34 prospective panellists who were then invited to participate. Of these, 19 (56%) responded to the first-round questionnaire, generating a panel falling within the 10 – 50 range of respondents regarded as ideal by policy Delphi methodologists. These 19 respondents were invited to respond to the second round and 14 accepted, an attrition rate of 26% but still producing a panel within the ideal range for deliberation. The profile of respondents in each of these two rounds is summarised in Table B1.

Table B1: Profile of panellists in first and second round of the policy Delphi.

Type of Expertise & Experience	No of panellists recruited		Pseudonyms for anonymous reporting	
	1st Round	2nd Round	1st Round	2nd Round
Non-Governmental Organisations concerned with facilitating collaboration between government, business and academia in addressing issues of food safety (NGO)	1	1	NGO1	NGO1
Digital commercial organisations concerned with promoting the uptake of digital technologies by government and business (DIGICOM)	2	1	DIGICOM1 DIGICOM2	DIGICOM2
Government food safety regulator (FSR)	2	1	FSR1 FSR2	FSR2
Food business organisations concerned with the safety of food	7	5	FBO1 FBO2	FBO2 FBO3

produce and its retail (FBO)			FBO3 FBO4 FBO5 FBO6 FBO7	FBO4 FBO6 FBO7
Academics involved in researching trends in food security (ACA)	7	6	ACA1 ACA2 ACA3 ACA4 ACA5 ACA6 ACA7	ACA1 ACA2 ACA3 ACA4 ACA5 ACA6
TOTAL	19	14		

This policy Delphi had two rounds. In keeping with the broader objectives of this report, panellists were asked the following semi-structured questions in the 1st round aimed at clarifying the principal issues about using advanced data analytics to enhance safety and authentication in the UK food and drinks sector:

> Question 1: In your opinion, what are the three biggest risks to food safety and authenticity in the next 2-5 years?
>
> Question 2: Are you aware of any examples where advanced data analytics has already helped to improve the management of any of the risks you have listed above?
>
> Question 3: In your opinion, in which 2-3 areas are advanced data analytics likely to have the biggest impact within the next 2-5 years on the management of food safety and authenticity risks?
>
> Question 4: In your opinion, what are the three key challenges facing the food sector for maximising the benefits of advanced data analytics in the management of food safety and authenticity risks?

As elaborated in the summary of key findings, answers to this 1st round identified seven prospective uses of advanced data analytics and five associated scenarios about their likely adoption.

In the 2nd round, a more structured questionnaire asked panellists to rank the technical and political feasibility of these prospective uses and sought their agreement with, predominantly sceptical, views on their likely adoption. Consequently, this deliberation helped to clarify some of the prospective challenges as well as opportunities for the adoption of advanced data analytics in this sector and these are elaborated in the following discussion of key findings from the policy Delphi.

Key Findings

Panellists identified six areas in which they thought advanced data analytics would be likely to have the biggest impact on food safety and authentication in the UK in the medium term:

1. creating a digital twin of the food system;
2. the use of global trade data;
3. development of the Food Hygiene Rating Scheme (FHRS);
4. combining diverse datasets;
5. use of Internet of Food Things (IoFT) sensors; and
6. automated detection of anomalies.

The 2nd round questionnaire presented respondents with a verbatim report of the responses from all panellists to the 1st round questionnaire and then asked them to score the technical and political feasibility of each of these uses on a Likert scale of: definitely feasible, probably feasible, may or may not be feasible, probably unfeasible, definitely unfeasible. Panellists were also invited to report the reasoning behind their rating in free text.

This produced some interesting clustering of opinion, including outliers that challenged the weight of agreement amongst other panellists. This is further illuminated by the free text reasoning provided by some panellists for their judgements.

1. Creation of a digital twin of the food system

Opinion about the technical feasibility of creating a digital twin of the food system was spread amongst three basic clusters of opinion and two outliers. One respondent who felt it was definitely feasible, argued:

"Technically this is possible and there are a number of projects where this is happening (digital city twin) or where this has been proposed (digital twin of Jersey.) This has been reliant on willing participants from across the system. For a roll-out then a data framework would need to be put in place to create confidence and maintain sensitivity." (FBO2)

Conversely, another panellist from the food business sector thought this proposal definitely unfeasible, arguing a digital twin would be:

"Impractical due to fragmented nature of the food system." (FBO7)

The remaining panellists clustered around those who thought it probably feasible, if the focus is upon specific sub-sectors rather than the food system in its entirety, and those who think it probably unfeasible, thinking it:

"Very hard to achieve with accurate results. Supply chain is fast moving and ever changing whilst being extremely complex." (FBO3)

Opinion about the political feasibility of this use of advanced data analytics was less certain still, thinking it may or may not be feasible depending on ‘buy-in’ amongst competitor businesses and across different jurisdictions.

2. Analysis of global trade data to anticipate food safety problems and traceability

Opinion on the technical feasibility of using global trade data to predict risky food-country combinations and improve the traceability of food commodities clustered around those thinking it probably, if not definitely, feasible although there is a need to further improve the quality, granularity and timeliness of this data. A more equivocal view argued that:

*“*The most serious risks are the ones that are hardest to predict.” (ACA5)

Opinion about the political feasibility of using global trade data in this way revealed a wider distribution of opinion, with one academic arguing it definitely feasible, the:

*“*Only barriers are the risk mitigations around predictions. Litigation?*”* (ACA6)

Others thinking it probably feasible noted the changing strategic context of global trade and the COVID-19 emergency which hit during the 2nd round collection of data for this research, a food standards regulator argued there are:

*“*No particular reasons to suspect public or stakeholder opposition but consideration of priorities may be necessary in the current COVID-19 operating environment*.”* (FSR2)

More sceptical panellists noted that, in this changing strategic context:

*“*This will be an interesting area of focus in a ‘post COVID-19’ environment where food security and traceability is key to public health. It is uncertain at this point how political attitudes will change with respect to traceability, but an opportunity clearly exists here.” (NGO1)

An academic panellist also noted that:

*“*Some countries might argue that algorithms are biased against them*.”* (ACA2)

Other academic panellists thought this usage probably unfeasible given the lack of trust around data sharing, the financial costs of accessing such data and the problem that such data might cultivate a false sense of security.

3. Developing FHRS into a more intelligence-led process

Again, opinion about the technical feasibility of this use of advanced data analytics clustered around those thinking it definitely or probably feasible with some arguing there is already the technology for predictive analytics using this and other imaginative sources of information, including social media:

*“*Availing of predictive analytics based on available public and official data would be an opportunity to focus resources on targeted inspections (e.g. data can be scraped from TripAdvisor or obtained from other regulatory agencies)*.”* (FBO4)

A more equivocal panellist questioned:

“Who’s intelligence? The most serious risks are the ones about which it will be hardest to obtain accurate information.” (ACA5)

This panellist reiterated this point in relation to the political feasibility of using the FHRS in this way. More equivocal panellists noted potential resistance to the collection of the data that would be needed, particularly the absence of a public education campaign about the reasons for this kind of surveillance:

“UK FSA and others have issues in collecting the Human Intelligence (HUMINT) to populate such a process. Food organisations would take a lot of persuasion and the public needs to be educated about what is required.” (ACA4)

“Political difficulties may stem from the transparency of any prioritised inspection regime (which currently will mainly be driven by what the last FHRS rating was).” (FSR2)

Even so, the weight of opinion on the panel clustered around the political feasibility of this suggestion, given the high regard and trust for the FHRS amongst good businesses and the general public:

“The FHRS in its current form is well known and appreciated by the wider public and its improvement is likely to be supported.” (ACA2)

“The Food hygiene rating scheme is widely accepted standard and system already developed. It has been implemented now long enough so that businesses where applicable have had time to address their system deficiencies. Technology would now be received as a means to support what is now perceived as a quality mark.” (NGO1)

4. Combining diverse datasets to improve quality and timeliness of predicting threats to food safety and authenticity

Opinion on the technical feasibility of combining data from administrative authorities, commercial sources and UGC, such as social media, satellite images and so forth, clustered around those who thought it definitely feasible, indeed already happening, and those thinking it probable, although with some reservations about the credence that can be given to some data sources:

“The accuracy and resolution of prediction is dependent on the quality and frequency of the data. Research is needed on data cleansing, especially with regard to social media.” (ACA6)

Those questioning the technical feasibility of this, noted problems in integrating diverse data sets:

“Unlikely to be successful without very granular high-quality data. Training data is limited and difficult to obtain. Relationships between risks and data are likely to be complex and difficult to model.” (ACA2)

“This is a very complex fusion exercise and the matching of signals of different shapes, sizes and significance will require a great deal of expert input into calibrating algorithms to do even some of the analysis or, alternatively, will continue to require human interpretation to complete the value add.” (FSR2)

One outlier opinion thought such data integration definitely unfeasible, arguing:

“Satellite images will not tell you want is happening under cover or in sheds, abattoirs or warehouses. Social media are notoriously unreliable.” (ACA5)

This panellist reiterated this point in arguing such data integration is definitely politically unfeasible. More equivocal panellists noted the challenge of sourcing key datasets needed for the corroboration of less robust sources such as social media:

"Issue is actually sourcing the data and turning it into meaningful and timely data. Most food businesses just want the 'top 5 I need to worry about this week'." (ACA4)

Even so, a majority of the panellists were more optimistic about the political feasibility of combining administrative, commercial and user-generated data sources:

"This should help a number of challenges that the food industry has today and needs coordinating at a national and if possible global level." (FBO2)

"No real political problem. The challenge would be around what level of investment could be justified and the need to show proof of concept." (FBO4)

5. Embedding Internet of Food Things (IoFT) sensors across food supply chain to detect micro-biological risks

Opinion on the technical feasibility of using IoFT sensors to detect micro-biological risks is distributed across three main clusters. Those thinking it definitely feasible argue the technology exists but there are barriers to adoption. The second cluster of opinion thinking this usage probably feasible, argue the technology is getting there albeit there is need for further research and development. This usage of IoFT is:

"More and more feasible as sensors become cheaper and more ubiquitous. However, not all the data will be accessible and that is the challenge. Pilot projects needed. However, taking the micro example, it could facilitate ecosystem modelling whereby the model can predict hazards." (FBO4)

"Again, technically feasible at a proof of concept level. Plenty of examples in research projects from around the world have existed for some time. Connecting monitoring technology with early warning systems technically possible. There would be technical (i.e. standards, API, interoperability) challenges in integrating systems assuming it was not one business delivering all of the elements. For a real world, i.e. trustworthy, solution a high degree of objective assurance would be needed, therefore necessitating a layer of validation across the systems to address risks such as tampering, untested devices, fraud etc." (ACA1)

More equivocal panellists identify uncertainty about the level of R&D that will be needed, the financial costs and incentives to develop, implement and monitor such sensors:

"May need incentivising as industry may be reluctant to invest. Needs to be linked to clear financial and other business benefits." (ACA6)

"Easy to do using existing technology and also already used but have scored to a 3 due to cost constraints." (FBO3)

Again, one panellist (ACA5) reiterated an objection to the feasibility of this use of the IoFT on both technical and political grounds:

"And who will ensure that they are not being interfered with?" (ACA5)

This, however, was very much an outlier opinion along with the more equivocal reasoning of another of the academic panellists:

“May be political and industry concerns about surveillance and scale of processes needed. Public may need convincing that sensors on food and packaging are not harmful.” (ACA4)

Otherwise the remaining, overwhelming majority, of the panel thought this use of the IoFT politically feasible:

“The political challenges, whilst achievable, would build upon the practical ones. Namely quality assurance, liability, responsibility, unintended consequences (if a product does not fail a check does that mean it is safe?). Who carries responsibility for the devices, the checks, the data, the modelling, etc.? That said, this is clearly something worth attempting, and learning from the process.” (ACA1)

“Given that this is most likely to be applied to food supply chains serving restaurant multiples, or big supermarkets, consumers might see benefit here (perhaps again dependent on how much of the cost migrates to them in end product prices).” (FSR2)

“The sensor technology, connectivity and analytics capability exist to enable the rollout of a IOFT strategy. Food production supply chains will require efficiency and early disruption signalling. Some challenges will exist on global supply chains but this very possible.” (NGO1)

6. Automated detection of anomalies in food products

Opinion about the technical feasibility of automating detection of anomalies in food products clustered around panellist type: FBOs, who thought it definitely feasible; academics, who thought it probably unfeasible; and a mixture of other panellists who were equivocal. The FBO advocates reasoned that automated detection was technically feasible because:

“There are multiple tests taken place across the value chain globally. Testing standards are in place. This needs to be a global database that is interoperable and can be integrated along the value chain.” (FBO2)

“Technically feasible using, e.g., full-scan LC-MS or fingerprinting technologies” (FBO4)

By contrast, the academic critics of automated detection argued:

“Just too many items in too many supply chains?” (ACA4)

“This will be problematic with many processed foods, the ingredients of which vary with input prices.” (ACA5)

More equivocal panellists acknowledged this as a likely future direction whilst acknowledging the technical complexity of the task:

“Whilst a good idea, that builds on great knowledge of components and testing processes, food is not pharma. Not all food is created in laboratory conditions. The ideas are good and definitely worth pursuing. Another significant risk is post-testing contamination. Therefore, all tests would have to be presented with caveats.” (ACA1)

“Lots of research in this area but very difficult. Spectroscopic techniques are promising but the challenge is the complexity of food products and performing measurements in non-laboratory environments.” (ACA3)

In the 2nd round, panellists were also asked to rate their agreement with five future scenarios in the use of advanced data analytics in the UK food and drinks sector over the medium term which, again, had been elicited from responses to the more open-ended 1st round questionnaire (see Table B2).

Other than the final scenario, regarding prospects for the automation of analysis of food safety and authenticity, a majority agreed with the forecasts. Although, even here, there were analytically significant clusters and outliers of opinion in disagreement with them. Further insight into this clustering of opinion was provided by those panellists who used the opportunity to record the reasoning behind their responses in free text.

Table B2: Scenarios for use of advanced data analytics in UK food and drink sector.

Scenario	Forecast
Analytical Skills	The supply of advanced data analytics skills for food safety and authenticity risk monitoring will continue to lag behind the demand, given education and training needs and the competition for these skills from multiple sectors, such as MedTech, FinTech and manufacturing.
Data Monetisation	Unless data on food safety can be monetised there will continue to be an insufficient supply of data of suitable quality for advanced analytics and thus of data-driven capabilities, as most food business organisations are low profit margin operations with a lack of sufficient resources to improve the capture and analysis of relevant information.
Risk Aversion	Advances in food safety and authenticity risk analytics will continue to be hampered by concern for the economic sensitivity of FBO data.
Predictive Modelling	Within the next five years, advanced data analytics will have access to collated, real-time, global datasets on, inter alia, commodity prices, international trade, adverse weather conditions, infectious disease outbreaks and geopolitical unrest, to predict effectively the risks of food safety and authenticity incidents.

Analytical Skills

This first, rather pessimistic, scenario forecast the continued lag of the requisite skills for data analysis in the service of improved food safety and authentication, given the education and training required to close the skills gap and given the intense competition for these analytical skills from wealthier and more lucrative sectors of the economy, especially FinTech, MedTech and manufacturing. The clustering of opinion was overwhelmingly in agreement with the likelihood of this forecast, with 86% agreeing:

“Research that I have been involved in has indicated that there is a gap between the skills that are needed and what is being provided by teaching and training establishments. This therefore puts a price premium on those with such skills, inevitable motivating practitioners towards fintech etc. Furthermore, the skillsets required are also evolving, so continuous learning is required. Therefore, career pathways with opportunities for access to lifelong learning will also prosper. Thirdly, data science practices also require interdisciplinary collaborative practices. From strategic leaders to those involved in implementing data capture solutions, as well as data stewards and visualisation experts all need to work together. So, it is as much about upskilling as recruitment.” (ACA1)

A small cluster was more equivocal, noting some grounds for optimism from a food safety regulator:

“This is a competitive skills area, but food can be an appealing area of business with broad relevance to the lives and lifestyles of all of us. Acknowledging the requirement for competitive salaries, or in government for the inclusion of this area of business within relevant high potential development streams, could increase the presence of these skills in this area.” (FSR2)

Data Monetisation

This, similarly pessimistic, scenario forecast a continued lack of data of sufficient quality to support the advanced analytics that could improve food safety and authentication unless a way can be found to monetise this data and thus provide its key providers, primarily food business organisations with limited additional resources to invest in data collection, with an incentive to invest in advanced data analytics. Here the clustering of opinion was more dispersed with half of the panel agreeing that this is likely to remain the case in the absence of clear commercial benefit or legal compulsion to collect and provide such data:

“Whether it is through the supply in return of actionable insight (leading to business and perhaps financial benefits) or in the demonstration of social responsibility (and the protection of public health and of consumers broader interests), data sharing will need either to be legally mandated or articulated as clearly good for business. It may also depend on the specific problem around which data is being collected (and the scale of enterprise from which the data is being requested).” (FSR2)

A significant minority (29%) of the panel disagreed with this forecast, arguing that:

“Food businesses can be incentivized in other ways for instance by increased consumer confidence. Also disagree that the problem is the availability or quality of data. There are currently many data sets with potential uses in food safety provided they are analysed/modelled properly.” (ACA2)

Risk Aversion

This scenario forecasts a situation in which advanced data analysis will be frustrated by the reasonable aversion of FBOs to the risk of undermining their market competitiveness. Here an interesting split occurred between an overwhelming majority (79%) agreeing with the forecast:

"True, however, such inclinations can be overcome by planning implementation strategies that incorporate some behavioural thinking to address and pre-empt such legitimate concerns." (ACA1)

"Fiscal and hygiene data on an owner of an FBO is very valuable to forecast and project future standards. Great sensitivity is required to persuade FBOs to share from fear of data leakage." (DIGICOM1)

No panellist equivocated but a small cluster disagreed with this forecast, arguing:

"The problem is not data availability it is the insufficient investment and competence in analysing and modelling currently available data sets." (ACA2)

"If value can be proven, and insight provided to drive improvement then this should make the system more efficient. There are examples of this happening already Food Industry Intelligence Network. This needs to be broader but is a case study to show it can work!" (FBO2)

Predictive Modelling

By contrast to the foregoing scenarios, this forecast expressed more optimistic responses on prospects for advanced data analytics that were elicited from responses to the 1st round questionnaire. It asked the 2nd round panellists to agree or disagree with the likelihood that within the next five years predictive modelling will have reached a level of sophistication capable of using real-time, global, datasets on, *inter alia*, commodity prices, trade, adverse weather conditions, infectious diseases, geopolitical unrest to effectively anticipate risks to food safety and authentication. There was a clear majority (71%) in agreement with this forecast. These more optimistic views argued:

"This is very probably true; however, the world is changing fast, some would say at an exponential rate. Therefore, models that are based on the past, even with this rich harvest of data, may not be capable of accurately predicating the future. Furthermore, even with such risk analysis – as with the current pandemic – would there be an appetite to fully invest in measures to fully mitigate the risk?" (ACA1)

"The data sets exist. Data layering is possible and predictive modelling is in practice. So, it should just be a case of applying the technologies to these data sets and growing the model." (FBO2)

"This is an ambitious but just about plausible prediction and I assess that there will be significant movement in this direction, perhaps even more so following a period of reflection and necessary innovation after COVID-19. My reservations expressed around item 5, however, may put a ceiling on the precision of this modelling." (FSR2)

"I have no evidence, but the pace of modelling is increasing. The key problem is not the modelling but the underlying datasets, the comprehensiveness and veracity of data, and the orderliness and cleanliness of the data." (ACA4)

Even so, a minority of panellists disagreed with this forecast, one arguing:

"The interactions of these factors are complex and hence very challenging to model mathematically. Even if such data were to become available acquisition, storage and processing of data is likely to be challenging." (ACA2)

Final Observations

Three panellists took the opportunity of a concluding free text question inviting any further, final, observations on prospects for the use of advanced data analytics in food safety and authentication. In addition to acknowledging the value of this deliberative exercise and a need to sustain the dialogue between government, commerce and academia in this field, these responses emphasised the centrality of developing datasets that are ‘inter-operable’, that can be meaningfully related to one another, and to the huge impact of the COVID-19 emergency which hit during the 2nd round of this policy Delphi. It is suggested the COVID-19 emergency will accelerate trends that were already under way, including the impact of eCommerce and home delivery on the governance and regulation of food standards, especially those associated with ‘Dark Kitchens’ (on-line food suppliers whose provenance is unclear and who are not open to inspection).

Bibliography

1. De Loë, R. C., Melnychuk, N., Murray, D., & Plummer, R. (2016). Advancing the state of policy Delphi practice: A systematic review evaluating methodological evolution, innovation, and opportunities. Technological Forecasting and Social Change, 104, 78-88.
2. Tendall, D. M., Joerin, J., Kopainsky, B., Edwards, P., Shreck, A., Le, Q. B., ... & Six, J. (2015). Food system resilience: defining the concept. Global Food Security, 6, 17-23.
3. Eggers, W. D., Turley, M., & Kishnani, P. (2018). The future of regulation: Principles for regulating emerging technologies. Deloitte center for government insights.
4. Jack, L. (2014). Food fraud is still hard to detect–so follow the money.
5. Bitzios, M., Lisa, J. A. C. K., Krzyzaniak, S. A., & Mark, X. U. (2017). Country-of-origin labelling, food traceability drivers and food fraud: Lessons from consumers' preferences and perceptions. European Journal of Risk Regulation, 8(3), 541-558.
6. Ingram, J., & Zurek, M. Food systems approaches for the future. Agriculture & Food Systems to, 2050, 547-568.
7. Yoo, Y., & Perzanowski, M. S. (2014). Allergic sensitization and the environment: latest update. Current allergy and asthma reports, 14(10), 465.
8. Changing consumption patterns: what about food safety? (2019). 15th Symposium of the Scientific Committee, Belgian Federal Agency for the Safety of the Food Chain.
9. Edizioni, C. (2012). Eating in 2030: trends and perspectives. Barilla Center for Food and Nutrition.
10. Morgan, O. (2019). How decision makers can use quantitative approaches to guide outbreak responses. Philosophical Transactions of the Royal Society B, 374(1776), 20180365.
11. Zoullouti, B., Amghar, M., & Nawal, S. (2019). Using Bayesian Networks for Risk Assessment in Healthcare System. In Bayesian Networks-Advances and Novel Applications. IntechOpen.
12. Pereira, J. C., Fragoso, M. D., & Todorov, M. G. (2016). Risk assessment using bayesian belief networks and analytic hierarchy process applicable to jet engine high pressure turbine assembly. IFAC-PapersOnLine, 49(12), 133-138.
13. Economou, T., Stephenson, D. B., Rougier, J. C., Neal, R. A., & Mylne, K. R. (2016). On the use of Bayesian decision theory for issuing natural hazard warnings. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 472(2194), 20160295.
14. Collazo, R. A., Görgen, C., & Smith, J. Q. (2018). Chain event graphs. CRC Press.
15. Mitchell, T. M. (1997). Machine learning. McGraw-Hill.
16. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
17. Elliott, A., Cucuringu, M., Luaces, M. M., Reidy, P., & Reinert, G. (2019). Anomaly detection in networks with application to financial transaction networks. arXiv preprint arXiv:1901.00402.
18. Chaters, G. L., Johnson, P. C. D., Cleaveland, S., Crispell, J., de Glanville, W. A., Doherty, T., ... & Salvador, L. C. M. (2019). Analysing livestock network data for infectious disease control: an argument for routine data collection in

emerging economies. Philosophical Transactions of the Royal Society B, 374(1776), 20180264.
19. Rushton, S. P., Sanderson, R. A., Reid, W. D., Shirley, M. D., Harris, J. P., Hunter, P. R., & O'Brien, S. J. (2019). Transmission routes of rare seasonal diseases: the case of norovirus infections. Philosophical Transactions of the Royal Society B, 374(1776), 20180267.
20. Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A. N., Gouws, S., ... & Sepassi, R. (2018). Tensor2tensor for neural machine translation. arXiv preprint arXiv:1803.07416.
21. Jiang, B., He, J., Yang, S., Fu, H., Li, T., Song, H., & He, D. (2019). Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues. Artificial Intelligence in Agriculture, 1, 1-8.
22. Zhang, Z., Geiger, J., Pohjalainen, J., Mousa, A. E. D., Jin, W., & Schuller, B. (2018). Deep learning for environmentally robust speech recognition: An overview of recent developments. ACM Transactions on Intelligent Systems and Technology (TIST), 9(5), 1-28.
23. Solangi, Y. A., Solangi, Z. A., Aarain, S., Abro, A., Mallah, G. A., & Shah, A. (2018). Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis. In 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS) (pp. 1-4). IEEE.
24. Chapman, A. B., Peterson, K. S., Alba, P. R., DuVall, S. L., & Patterson, O. V. (2018). Hybrid system for adverse drug event detection. In International Workshop on Medication and Adverse Drug Event Detection (pp. 16-24).
25. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. ACM Computing Surveys (CSUR), 51(2), 1-36.
26. Kulkarni, A., & Shivananda, A. (2019). Deep Learning for NLP. In Natural Language Processing Recipes (pp. 185-227). Apress, Berkeley, CA.
27. Khatua, A., Khatua, A., & Cambria, E. (2019). A tale of two epidemics: Contextual Word2Vec for classifying twitter streams during outbreaks. Information Processing & Management, 56(1), 247-257.
28. Newbold, B., Watson, S., Mackay, K., & Isaacs, S. (2013). Exploring the Relationship between Food Access and Foodborne Illness by Using Spatial Analysis. *Journal of food protection*, *76*(9), 1615-1620.
29. Chaikaew, N., Tripathi, N. K., & Souris, M. (2009). Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand. *International Journal of Health Geographics*, *8*(1), 36.
30. Spencer, S. E. F., Marshall, J., Pirie, R., Campbell, D., Baker, M. G., & French, N. P. (2012). The spatial and temporal determinants of campylobacteriosis notifications in New Zealand, 2001–2007. *Epidemiology & Infection*, *140*(9), 1663-1677.
31. https://rspatial.org/raster/analysis/3-spauto.html
32. https://gisgeography.com/kriging-interpolation-prediction/
33. Gottwald, T., Luo, W., Posny, D., Riley, T., & Louws, F. (2019). A probabilistic census-travel model to predict introduction sites of exotic plant, animal and human pathogens. Philosophical Transactions of the Royal Society B, 374(1776), 20180260.
34. https://www.nao.org.uk/press-release/ensuring-food-safety-and-standards/

35. Guiñazú, M. F., Cortés, V., Ibáñez, C. F., & Velásquez, J. D. (2020). Employing online social networks in precision-medicine approach using information fusion predictive model to improve substance use surveillance: A lesson from Twitter and marijuana consumption. Information Fusion, 55, 150-163.
36. Jayawardhana, U. K., & Gorsevski, P. V. (2019). An ontology-based framework for extracting spatio-temporal influenza data using Twitter. International journal of digital earth, 12(1), 2-24.
37. Chaix, E., Deléger, L., Bossy, R., & Nédellec, C. (2019). Text mining tools for extracting information about microbial biodiversity in food. Food microbiology, 81, 63-75.
38. Mcloughlin, MP., Stewart, R., & McElligott, A. (2019) Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. Journal of the Royal Society Interface 16.155: 20190225.
39. Halachmi, I., Guarino, M., Bewley, J., & Pastell, M. (2019). Smart animal agriculture: application of real-time sensors to improve animal well-being and production. Annual review of animal biosciences, 7, 403-425.
40. Stern, U., He, R., & Yang, C. H. (2015). Analyzing animal behavior via classifying each video frame using convolutional neural networks. Scientific reports, 5, 14351.
41. Arac, A., Zhao, P., Dobkin, B. H., Carmichael, S. T., & Golshani, P. (2019). DeepBehavior: A deep learning toolbox for automated analysis of animal and human behavior imaging data. Frontiers in systems neuroscience, 13, 20.
42. Miller, G. A., Hyslop, J. J., Barclay, D., Edwards, A., Thompson, W., & Duthie, C. A. (2019). Using 3D imaging and machine learning to predict liveweight and carcass characteristics of live finishing beef cattle. Frontiers in Sustainable Food Systems, 3, 30.
43. Verhaelen, K., Bauer, A., Günther, F., Müller, B., Nist, M., Celik, B. Ü., ... & Wallner, P. (2018). Anticipation of food safety and fraud issues: ISAR-A new screening tool to monitor food prices and commodity flows. Food Control, 94, 93-101.
44. Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., & McFarlane, D. (2019). Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. International Journal of Production Research, 1-12.
45. Deleris, L. A., Elkins, D., & Paté-Cornell, M. E. (2004). Analyzing losses from hazard exposure: a conservative probabilistic estimate using supply chain risk simulation. In Proceedings of the 2004 Winter Simulation Conference, 2004. (Vol. 2, pp. 1384-1391). IEEE.
46. Wegner, A. E., Ospina-Forero, L., Gaunt, R. E., Deane, C. M., & Reinert, G. (2018). Identifying networks with common organizational principles. *Journal of Complex Networks*, *6*(6), 887-913.
47. Burkholz, R., & Schweitzer, F. (2019). International crop trade networks: the impact of shocks and cascades. *Environmental Research Letters*, *14*(11), 114013.
48. Garre, A., Fernandez, P. S., Brereton, P., Elliott, C., & Mojtahed, V. (2019). The use of trade data to predict the source and spread of food safety outbreaks: An innovative mathematical modelling approach. Food Research International, 123, 712-721.
49. Pantazi, X. E., Tamouridou, A. A., Alexandridis, T. K., Lagopodi, A. L., Kontouris, G., & Moshou, D. (2017). Detection of Silybum marianum infection

with Microbotryum silybum using VNIR field spectroscopy. Computers and Electronics in Agriculture, 137, 130-137.
50. Pantazi, X. E., Moshou, D., Oberti, R., West, J., Mouazen, A. M., & Bochtis, D. (2017). Detection of biotic and abiotic stresses in crops by using hierarchical self organizing classifiers. Precision Agriculture, 18(3), 383-393.
51. Ebrahimi, M. A., Khoshtaghaza, M. H., Minaei, S., & Jamshidi, B. (2017). Vision-based pest detection based on SVM classification method. Computers and Electronics in Agriculture, 137, 52-58.
52. Chung, C. L., Huang, K. J., Chen, S. Y., Lai, M. H., Chen, Y. C., & Kuo, Y. F. (2016). Detecting Bakanae disease in rice seedlings by machine vision. Computers and electronics in agriculture, 121, 404-411.
53. Ferentinos, K P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture 145: 311-318.
54. Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. Computers and Electronics in Agriculture, 121, 57-65.
55. https://www.oceanmind.global/
56. Brown, E., Dessai, U., McGarry, S., & Gerner-Smidt, P. (2019). Use of whole-genome sequencing for food safety and public health in the united states. *Foodborne pathogens and disease*, *16*(7), 441-450.
57. https://www.food.gov.uk/sites/default/files/media/document/csa-whole-genome-seq-reportv2%20(1).pdf
58. Wray, Naomi R., et al. (2019). Complex trait prediction from genome data: contrasting EBV in livestock to PRS in humans: genomic prediction. Genetics 211.4: 1131-1141.
59. Jarvis, K. G., Daquigan, N., White, J. R., Morin, P. M., Howard, L. M., Manetas, J. E., ... & Grim, C. J. (2018). Microbiomes associated with foods from plant and animal sources. Frontiers in microbiology, 9, 2540.
60. Brown, E., Dessai, U., McGarry, S., & Gerner-Smidt, P. (2019). Use of whole-genome sequencing for food safety and public health in the united states. Foodborne pathogens and disease, 16(7), 441-450.
61. Wheeler, N. E., Gardner, P. P., & Barquist, L. (2018). Machine learning identifies signatures of host adaptation in the bacterial pathogen Salmonella enterica. PLoS genetics, 14(5).
62. Deelder, W., Christakoudi, S., Phelan, J., Diez Benavente, E., Campino, S., McNerney, R., ... & Clark, T. G. (2019). Machine learning predicts accurately Mycobacterium tuberculosis drug resistance from whole genome sequencing data. Frontiers in Genetics, 10, 922.
63. Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., ... & McClure, P. (2019). The use of next generation sequencing for improving food safety: Translation into practice. Food microbiology, 79, 96-115.
64. https://www.efsa.europa.eu/en/supporting/pub/en-1154
65. http://www.fao.org/ag/againfo/programmes/en/empres/home.asp
66. Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics, 55(4), 665-676.
67. Höhle, M., & an der Heiden, M. (2014). Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011. *Biometrics*, *70*(4), 993-1002.
68. Park, Y., Kim, M., Pachepsky, Y., Choi, S. H., Cho, J. G., Jeon, J., & Cho, K. H. (2018). Development of a nowcasting system using machine learning

approaches to predict fecal contamination levels at recreational beaches in Korea. *Journal of environmental quality*, *47*(5), 1094-1102.
69. CSA report, issue eight: data standards, April 2019.
70. Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
71. https://www.turing.ac.uk/research/research-projects/artificial-intelligence-data-analytics-aida
72. Trunfio, P., & Vlassov, V. (2019). Clouds for scalable Big Data processing. International Journal of Parallel, Emergent and Distributed Systems.
73. Pearson, S., May, D., Leontidis, G., Swainson, M., Brewer, S., Bidaut, L., ... & Zisman, A. (2019). Are Distributed Ledger Technologies the panacea for food traceability?. Global food security, 20, 145-149.
74. Zhang, C., Gong, Y., Brown, S., & Li, Z. (2019). A content based literature review on the application of blockchain in food supply chain management.
75. Yu, F. R., Liu, J., He, Y., Si, P., & Zhang, Y. (2018). Virtualization for distributed ledger technology (vDLT). *IEEE Access*, *6*, 25019-25028.
76. https://theodi.org/article/odi-data-trusts-report/
77. Leslie, D. (2019). Understanding artificial intelligence ethics and safety. A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529
78. Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. Proceedings of FAT.
79. Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., ... & Eckersley, P. (2020). Explainable machine learning in deployment. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.
80. Global AI Race. AI in UK: Artificial Intelligence Industry Landscape Overview. Deep Knowledge Analytics (2018). http://analytics.dkv.global/data/pdf/AI-in-UK/AI-in-UK-Section-III-Global-AI-Race.pdf
81. Wang, W., Ning, Y., Rangwala, H., & Ramakrishnan, N. (2016). A multiple instance learning framework for identifying key sentences and detecting events. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (pp. 509-518).
82. Lampos, V., & Cristianini, N. (2012). Nowcasting events from the social web with statistical learning. ACM Transactions on Intelligent Systems and Technology (TIST), 3(4), 1-22.
83. Thomas, A., & Sangeetha, S. (2020). Deep Learning Architectures for Named Entity Recognition: A Survey. In Advanced Computing and Intelligent Engineering (pp. 215-225). Springer, Singapore.
84. Middleton, S. E., Kordopatis-Zilos, G., Papadopoulos, S., & Kompatsiaris, Y. (2018). Location extraction from social media: Geoparsing, location disambiguation, and geotagging. ACM Transactions on Information Systems (TOIS), 36(4), 1-27.
85. Köckerling, E., Karrasch, L., Schweitzer, A., Razum, O., & Krause, G. (2017). Public Health Research Resulting from One of the world’s Largest Outbreaks Caused by entero-Hemorrhagic Escherichia coli in Germany 2011: A Review. Frontiers in public health, 5, 332.
86. Mcloughlin, MP., Stewart, R., & McElligott, A. (2019) Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. Journal of the Royal Society Interface 16.155: 20190225.

87. Halachmi, I., Guarino, M., Bewley, J., & Pastell, M. (2019). Smart animal agriculture: application of real-time sensors to improve animal well-being and production. Annual review of animal biosciences, 7, 403-425.
88. https://datasciencecampus.ons.gov.uk/projects/faster-indicators-of-uk-economic-activity-shipping/
89. Arslanalp, M. S., Marini, M. M., & Tumbarello, M. P. (2019). Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time. International Monetary Fund.
90. Retkute, R., Jewell, C. P., Van Boeckel, T. P., Zhang, G., Xiao, X., Thanapongtharm, W., ... & Tildesley, M. J. (2018). Dynamics of the 2004 avian influenza H5N1 outbreak in Thailand: The role of duck farming, sequential model fitting and control. Preventive veterinary medicine, 159, 171-181.
91. Probert, W. J., Jewell, C. P., Werkman, M., Fonnesbeck, C. J., Goto, Y., Runge, M. C., ... & Tildesley, M. J. (2018). Real-time decision-making during emergency disease outbreaks. PLoS computational biology, 14(7), e1006202.
92. Manitz, J., Kneib, T., Schlather, M., Helbing, D., & Brockmann, D. (2014). Origin detection during food-borne disease outbreaks-a case study of the 2011 ehec/hus outbreak in Germany. PLoS currents, 6.
93. Garre, A., Fernandez, P. S., Brereton, P., Elliott, C., & Mojtahed, V. (2019). The use of trade data to predict the source and spread of food safety outbreaks: An innovative mathematical modelling approach. Food Research International, 123, 712-721.
94. Ji, G., & Tan, K. (2017). A Big Data Decision-making Mechanism for Food Supply Chain. In MATEC Web of Conferences (Vol. 100, p. 02048). EDP Sciences.
95. Njage, P. M. K., Leekitcharoenphon, P., & Hald, T. (2019). Improving hazard characterization in microbial risk assessment using next generation sequencing data and machine learning: predicting clinical outcomes in shigatoxigenic Escherichia coli. International journal of food microbiology, 292, 72-82.
96. Zhang, S., Li, S., Gu, W., den Bakker, H., Boxrud, D., Taylor, A., ... & Brown, E. (2019). Zoonotic source attribution of Salmonella enterica serotype Typhimurium using genomic surveillance data, United States. Emerging infectious diseases, 25(1), 82.
97. Vilne, B., Meistere, I., Grāntiņa-Ieviņa, L., & Ķibilds, J. (2019). Machine Learning Approaches for Epidemiological Investigations of Food-Borne Disease Outbreaks. Frontiers in microbiology, 10, 1722.