Chief Scientific Adviser's
Science Report

Issue six: Data Science

Food Standards Agency
food.gov.uk

"In my sixth Chief Scientific Adviser's Report, we look at an area that is fast transforming our lives – data science. Though it is fair to say that data science is used across all areas of government, its use is being increasingly realised within the regulation of food safety. From using sensors to check the temperature of the food we eat, to using big data techniques to help us build a better picture of the hygiene practices of food businesses, data science is essential in enabling the FSA to ensure the food you eat is safe to eat."

Professor Guy Poppy,
FSA Chief Scientific Adviser

The report shows how the department is bringing together a range of data – such as economic conditions, social media and consumer preferences – to help the FSA meet its statutory obligations. This includes supporting one of the FSA's key strategies moving forward, 'Regulating our Future', a wide-reaching change programme that will modernise the regulation of food within England, Wales and Northern Ireland. Specifically, the report looks at several case studies, which show how social media can be effectively harnessed. Our award-winning work on predicting norovirus outbreaks through tweets, and the analysis of health claims regarding raw milk on YouTube are two such examples.

In this constantly evolving field, developing partnerships with both producers and users of data is essential. This is especially pertinent with regard the developing field of big data science. That is why I'm proud to present some of the key partnerships that the FSA has formed, doing exciting research with University College London's Big Data Institute, the Office for National Statistics, and the Government Digital Service, to name a few.

All of this work we are doing in the field of data science will help us achieve a core FSA objective, 'to become a data-driven organisation', through realising the benefit of using data assets, both from within and outside of the department.

Finally, one area that can greatly benefit from the use of data science is surveillance. This report will show why surveillance is of importance to the FSA, and how big data feeds into the implementation of our new and innovative surveillance strategy.

Introduction

Food is an incredibly important part of our lives: everyone has to eat and drink to survive. It is, therefore, not surprising that the food industry is one of the largest and most important industries in the world. The food industry encompasses everything from food production and shipping companies to supermarkets and restaurants, but also includes food regulators like the FSA. This complex food 'ecosystem' generates an enormous amount of data, such as information about consumer purchase histories, allergy and nutrition information, and data from premise inspections including food hygiene inspections.

Data is fast transforming our lives: complex algorithms on retail websites seem to know exactly what we want based on previous purchases, sometimes suggesting items we have forgotten in our online shop. Data science approaches, such as predictive analytics and machine learning[1], are paying huge dividends for many companies by using their data to generate novel insights to increase their revenue and efficiency. But how does it apply to food regulation and food safety?

The data pyramid
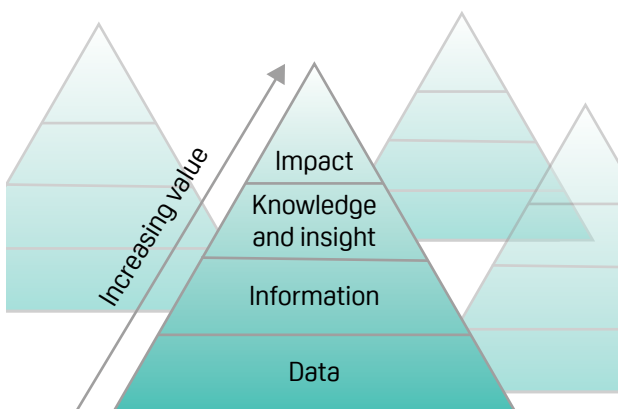
The data pyramid shows how we use data to create impact. First, raw data lay the foundations for impact, so it is essential that the 'right' data is included in addition to a good breadth of data. Ideally, this data already exists as collecting data is often expensive and time consuming. The value of data is most readily realised high in the pyramid (knowledge and impact) and by making our data open, and making effective use of others' data through a collaborative approach, we strive to maximise value and impact.

Insights are used to direct action, which in turn leads to impact.

Data analysis extracts information from these data and adds context; expert advice and engagement with others provides us with knowledge and insight from this information. These insights are then used to direct action, which in turn leads to impact - including the protection of consumers. Data science approaches allow us - more than ever before – to make smarter use of data to create this impact.

This report shows how we use data to fulfill our regulatory duties - it shows how our application of data science approaches and use of social media data leads to insight and impact, and it explains how we work with other Government departments and academia to build capability and to deliver best value from our data.

---

1. Machine learning gives computers the ability to learn without being explicitly programmed. Predictive analytics is an analytical technique that uses data to predict trends and patterns that will occur in the future.

Regulating our Future: redesigning the FSA's regulatory role

What data science opportunities are there for a regulator tasked to ensure that 'Food is safe' and that 'Food is what it says it is'? In a nutshell, data enables us to take advantage of new developments to ensure that food is safe. Specifically, new cost-effective technology (eg sensors) can continuously and automatically monitor food temperatures, where food safety inspections only provide a snapshot of an establishment's condition and compliance with our safety regulations at a single point in time. Clearly, the data generated through these technologies provide much greater insight than what could have been obtained through inspections or audits. This allows close to real time monitoring, which in turn allows the food business, and if necessary Food Standards Agency, to act immediately.

Data enables us to take advantage of new developments to ensure that food is safe.

REGULATING OUR FUTURE

'Regulating our Future' is an ambitious change programme that will tap into these innovative technologies including by auditing data provided by food businesses (eg from fridge thermometers or thermometers on packaging) to track compliance with necessary standards.

Likewise, we have ambition for data to transform our surveillance approach, bringing together a wide range of data (eg economic conditions, sensor data, consumer purchase data) from a variety of sources (eg reports from consumers, industry, trends on social media) to produce a comprehensive view of emerging risks in the food/feed chain (see the 'also of interest' section for more information).

A data-driven Food Standards Agency[2]

We have a wealth of data at our fingertips in the FSA, ranging from data about meat hygiene inspections, to data about hygiene in food premises. These data, usually collected as part of regulatory processes, form the basis for reports to our board and underpin public information such as the Food Hygiene Ratings Scheme[3].

Data is critical to achieving the Agency's objectives and is likely to increase in importance. In 2016/2017 the Agency embarked on a corporate objective 'to become a data-driven organisation', realising that we benefit from making more effective use of all data assets, both the Agency's and others, to support the delivery of our objectives. While we are working to develop and deliver a data driven world, we are also taking forward more traditional approaches to using data, such as driving improvements in performance, interventions and engagement.

Data is critical to achieving the Agency's objectives.

---

2. What does it mean to be data-driven? That the Agency uses data to: guide decision making, set objectives, standards and targets, prioritise work and monitor progress.

3. The Food Hygiene Rating Scheme (FHRS) in England, Wales and Northern Ireland helps consumers choose where to eat out or shop for food by giving them information about the hygiene standards in restaurants, takeaways and food shops.

Network of interactions between coworkers
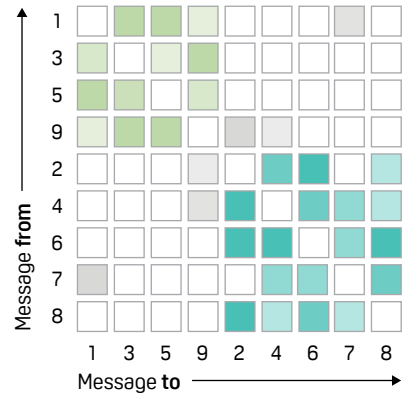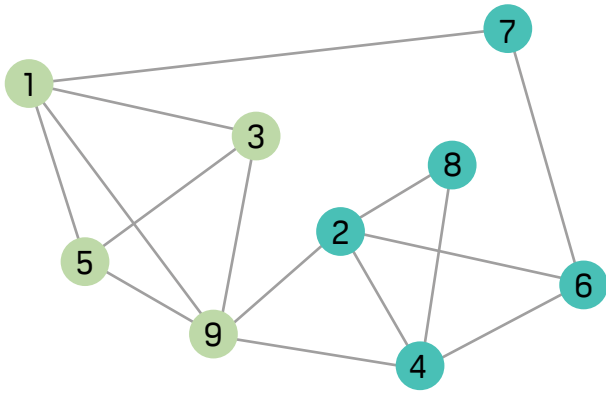
Community A   Community B   Weaker interaction   Stronger interaction

An example of how we analyse communities within our Yammer: the left panel shows which (fictional) co-workers have interacted shown by a line between them, the right panel shows the strength of interaction between co-workers, where a darker colour shows there has been more interaction between them. Different colours show different communities of co-workers, with gray squares showing interactions outside their own community.

One such project is using dashboards across the FSA, so that people can share interactive visualisations to explore data.

An early example of our work on dashboards is the analysis of our internal social network (Yammer) that was carried out to inform our internal engagement strategy. The dashboard shows how and when our employees engage on Yammer, identifying distinct communities of users through network analyses.

We are using dashboards across the agency, for example to help us plan and prioritise investments – where it helps those responsible for decision making see the impact of the investment in relation to the overall expenditure. Similarly, using external (publicly available) data, we created a dashboard for our food imports team that allows them to visualise the imports to and exports from the UK and to detect anomalies in trade, as well as identifying new areas of risk and vulnerability.

CASE STUDIES:

An early warning tool to predict norovirus outbreaks through tweets

Norovirus, also known as the winter vomiting bug, is the most common cause of gastroenteritis in the UK. It causes vomiting and diarrhoea with symptoms usually lasting around one to two days. Norovirus is highly contagious, and outbreaks frequently cause major disruption to essential services in places where a lot of people gather (eg schools, hospitals and care homes), with an estimated cost to the NHS of £100 million per year.

Inspired by early success of Google's Flu Trends, the FSA developed a predictive algorithm that compared social media data with foodborne illness statistics to provide early warnings of norovirus outbreaks. The Agency tracks the weekly number of tweets that contain specific keywords that indicate someone may have norovirus and correlate these numbers with the official statistics on confirmed norovirus cases from Public Health England. We then exclude tweets that are clearly not related to norovirus because they contain irrelevant keywords, for example tweets related to pregnancy (morning sickness), excessive

Diarrhoea and vomiting?

There's no specific cure for stomach bugs such as Norovirus
Treat yourself at home with an #EssentialKit while the virus runs its course

stay hydrated — take paracetamol — prevent spread — stay at home for two days after symptoms clear

alcohol consumption and other things that might justify the inclusion of 'sick' or similar words.

There is a lag of about two weeks between people getting ill and lab results being published, with historic data allowing us to create a predictive model. A keen eye is kept on these predictions when a significant rise in the number of cases is predicted: when there are three consecutive weeks of significant rises in the predicted cases of norovirus, confidence that the data is real is such that an intervention is triggered. That means action is taken at an earlier time than would be the case if laboratory reports were the instigator.

Number of tweets compared to confirmed lab reports

Number of tweets —— Number of confirmed lab reports
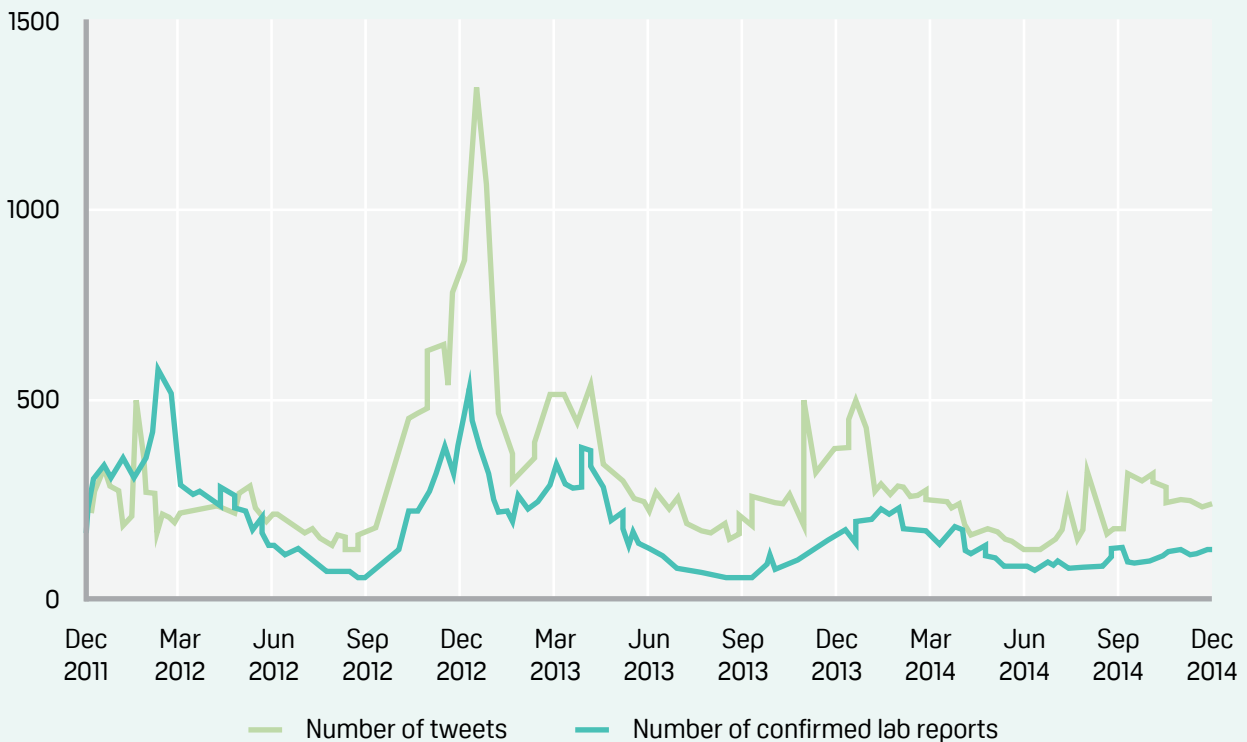
When we do detect an outbreak, we work with NHS Choices to limit the disruptive effect of an outbreak on essential services by rolling out interventions. Since there is no known cure for norovirus, the intervention materials – such as posting infographics on social media – are designed to make people aware of how to stop the spread of the virus (eg by washing your hands and staying at home).

We work with NHS Choices to limit the disruptive effect of an outbreak.

Earlier this winter we predicted an outbreak of norovirus, the first predicted outbreak since we started using the tool.

These are deployed to social media, but also to GP practices and schools, thereby reaching as wide an audience as possible. Earlier this winter we predicted an outbreak of norovirus, the first predicted outbreak since we started using the tool. We are carrying out a process and impact evaluation of effectiveness of the approach.

Emerging food trends on social media

Anne Burns joined the FSA from the University of Sheffield in the autumn, on a 6 month secondment to look at analysis of social media images. During her secondment Anne has looked at analysing Youtube content using raw drinking milk as an example, as well as considering the different audiences and publishers on Instagram and Pinterest through Food Trucks.

Anne has used image analysis to determine whether images from social media (eg Instagram) can provide insight into problematic food preparation practices, from undercooking to overcooking. Below are two studies that Anne has completed to gain insight into consumer attitudes and the latest food trends, and help to tailor the FSA's social media campaigns in a way that responds to particular online conversations and audiences.

Raw milk

Unpasteurised, or 'raw', milk may contain harmful bacteria that can cause food poisoning, therefore vulnerable groups, such as pregnant women and children, are advised not to consume it. However, some people believe that drinking raw milk is beneficial, so this case study sought to understand what these assumed benefits might be whilst investigating the suitability of image data on the platform as a means of gathering information and evidence.

Unpasteurised, or 'raw', milk may contain harmful bacteria that can cause food poisoning.

Raw milk health claims
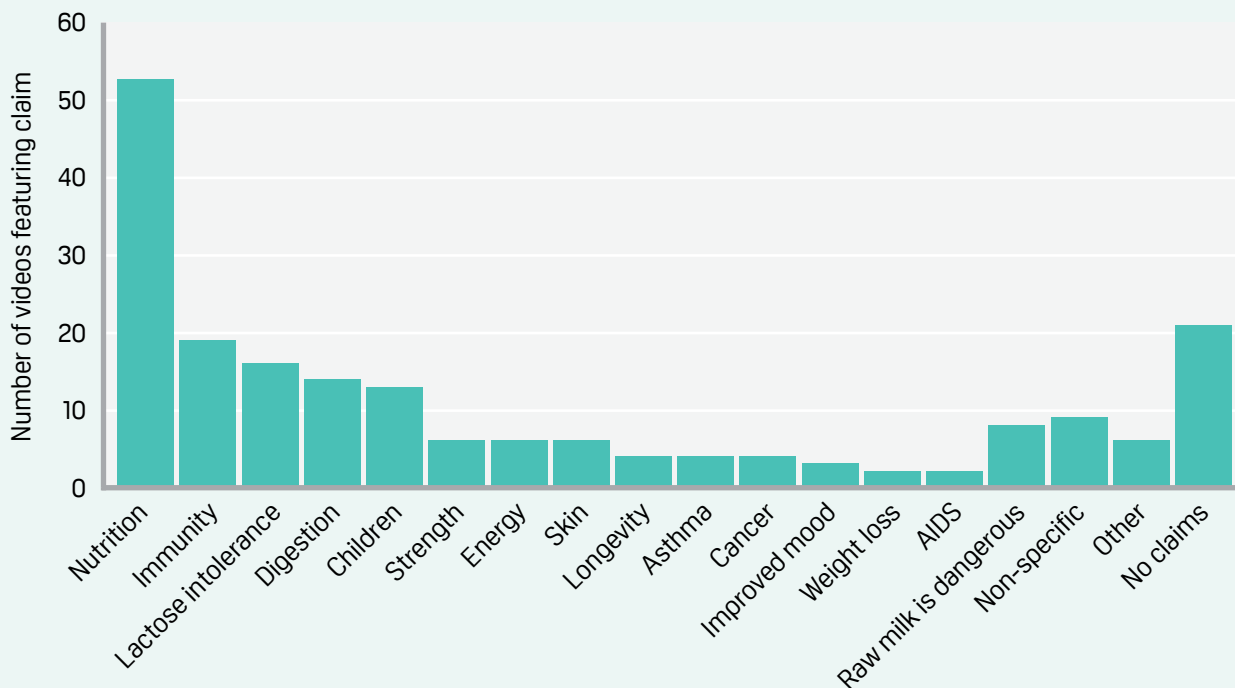
The 100 most viewed videos on YouTube relating to raw milk were watched and their content was noted. 92% of the videos supported raw milk, citing the value of natural or traditional foods, and the importance of the right to choose. Supporters argued that raw milk could help with lactose intolerance and improve immunity. These claims are not supported by medical research and some videos were particularly concerning, as they suggested giving raw milk to children and babies, or asserted that it could cure diseases such as cancer or AIDS. This kind of misinformation, in which personal opinion contradicts expert advice highlights the importance of understanding the ranges of evidences presented online.

Supporters argued that raw milk could help with lactose intolerance and improve immunity.

Food trucks

This case study looked at 600 images from Instagram and from Pinterest, in order to understand the online conversation concerning food trucks and to compare the data contained within these two image hosting platforms.

Pinterest featured mostly images of trucks (86% of images), whereas Instagram featured more images of food (55% of images, in comparison with Pinterest's 9%). Additionally, around 75% of the images on both sites had been either taken or shared by a business. The difference was in terms of what was being sold: on Pinterest, links to businesses selling trucks suggested a focus on food truck vendors, whereas on Instagram, an emphasis on depicting menus, opening times and locations suggested a conversation with customers. Therefore, both sites offer a useful opportunity for the FSA to communicate: Pinterest with mobile food caterers and Instagram directly with consumers.
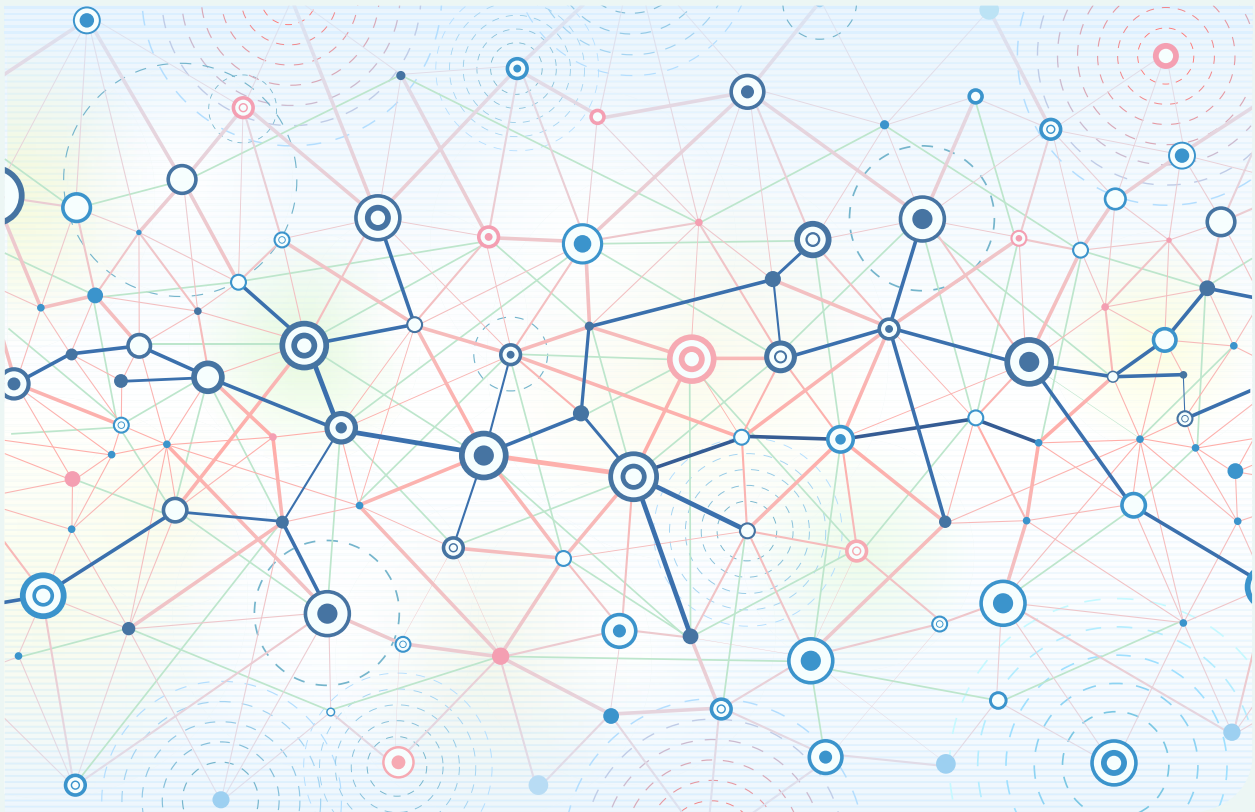
COLLABORATIONS

Data science and Food research fellowship
University College London's Big Data Institute

This year marks the start of an exciting collaboration between the FSA and University College London's Big Data Institute (UCL BDI). The FSA is funding a research fellow in Data Science and Food. Of course, food and food safety are at the heart of this collaboration, which will focus on developing new big data methods to understand risk and contagion foodborne illnesses, and other food safety issues.

Dr. Anatol Wegner - who has recently been recruited to the position - says "We will focus on developing new mathematical methods for analysing large collections of noisy and incomplete data sets with a strong emphasis on food science applications. Initially we will focus on methods that help find patterns in data related to the emergence and spread of foodborne illness outbreaks. Having access to the FSA's data and expertise presents a unique opportunity to develop research that has immediate impact on the ground."

The Data Science Campus
The Office for National Statistics

The Data Science Campus at the Office for National Statistics (ONS) is a hub and incubator for Data Science for Public Good, working collaboratively with the broader data science community to build impactful, co-creative and high value partnerships across the data landscape.

Based in Newport, South Wales, the campus brings together analysts, data scientists and technologists from across the UK and the wider international community. It acts as a hub for the whole of the UK public and private sectors to gain practical advantage from increased investment in data science research and help cement the UK's reputation as an international leader in this field. By partnering with other areas of government, industry and academia, ONS is developing a greatly enhanced range of measures of the economy and society, so that emerging issues and trends can be spotted more quickly and understood in greater detail, so that decision making is better informed.

While much of the research conducted at the Data Science Campus is focused on economic statistics, a broad spectrum of topics are explored as approaches developed can then be used to address a multitude of challenges. For example, the ONS gathers and analyses a wide range of data on UK dietary habits. The Living Costs and Food Survey, and the National Diet & Nutrition Survey suggest that calorie consumption has declined over the last 40 years. However reported levels of obesity and national population weight are increasing. The Data Science Campus is currently analysing the delta between these two indices. The big data and data science techniques developed in this research will have a wide range of applications across ONS and wider government.

The ONS gathers and analyses a wide range of data on UK dietary habits.

The ultimate goal of the Data Science Campus is to build this new generation of tools and technologies to exploit the growth and availability of innovative data sources and to provide rich informed measurement and analyses on the economy, the global environment and wider society.

The future

Big data and data science bring relatively new tools and techniques to Government analytics - tools that are opening a new chapter for us to learn, to take actions, and to create impact and create evidence-based policies. We are starting to explore these approaches more and more in our work so that we can learn more from our data. We also share our data and openly use data from different sources, collected by different people and organisations (eg in the 'Regulate our Future' programme) to maximise the potential for impact and keep the regulatory burden to a minimum.

Department for Envi l Affai

Government Digital Service

Office for National Statistics

WARWICK THE UNIVERSITY OF WARWICK

The University Of Sheffield.

UCL

Big data and data science bring relatively new tools and techniques to Government analytics.

Opportunities for data science in food safety and standards – beyond what we discuss in this report – are abundant, but generally we only use a small amount of the relevant data that is generated and available. It is clear that there is a need for a multidisciplinary approach across industry, government and academia to develop expertise and agree common approaches and standards. This will develop more impact from the available data and also might encourage data-rich organisations to make data available for use by others.

The challenges range from relatively simple, such as switching from handwritten reports to electronic reports and making data openly available in formats that support reuse, to more complex, such as building an infrastructure that allows us to check compliance with our regulatory requirements from data collected by food businesses.

An important next step will be to develop the capacity and capability that enables us to make more effective use of data to identify and act upon food safety issues. To realise this ambition, the agency continues to work with food businesses, academia (eg University College London, University of Sheffield, Warwick University) and other Government departments (The Office for National Statistics, the Government Digital Service, and the Department for the Environment, Food and Rural Affairs) to get the best value from data to support our objectives and learn with and from others.

Acknowledgements

Also of interest: The FSA Surveillance Strategy

What is surveillance and why is it so important to the FSA?

The World Health Organisation defines surveillance as the 'ongoing systematic collection, collation, analysis and interpretation of data and the dissemination of information to those who need to know in order that action may be taken'. For the FSA, surveillance involves the collection of a wide range of data that can inform us about the food system. The resulting insights can then be used to manage risks and seize opportunities by not only ourselves, but by others such as the food/feed industries, enforcers, consumers, etc.

Surveillance is fundamental to the work of the FSA and is integral to ensuring the delivery of our strategic outcomes, to inform future targeted official control activities and to define future strategic priorities.

Where are we now?

The UK food surveillance system is a constantly evolving and complex process involving a vast array of intelligence and sources of data. Every division within the FSA is involved in food surveillance in one form or other.

FSA current activity includes, but is not limited to, collection of data for which we are the enforcing authority (such as in abattoirs), scanning the horizon for potential future risks and collecting and analysing intelligence in the fight against food crime.

The FSA does not work in isolation in this field – it works as part of a surveillance network. We collaborate with other Government Departments, the food industry and its trade bodies, local enforcement authorities, consumers, regulators and laboratories to ensure our surveillance activities are relevant and important to our goal of ensuring food safety. The FSA surveillance network is extensive and far-reaching, involving the complex flow of data between all those involved.

The diagram on the next page shows some of the main players involved in the UK food surveillance system.

What is the future?

The food surveillance system is constantly changing and is affected by new foods and food sources, new technologies and new businesses. The FSA is proactive in evolving our surveillance approach. We are looking at historical information, learning lessons from this, and utilising it to shape our future goals and aspirations.

The FSA is currently in an exciting phase in the development of a new strategic approach to surveillance. The ultimate aim of this new approach is to develop a flexible evolving surveillance system to account for possible risks and challenges to the food/feed system. This will involve us strengthening our internal and external surveillance networks to continue the effective collaboration in this field.

The surveillance strategy will also ensure that we have effective utilisation of Big Data, to forecast trends and generate information and insight from large data sets, allowing us and others to identify and manage emerging risks for the future.

Our new approach will offer many benefits to all those involved in the food surveillance network, not only the FSA, including creative problem solving, maximising the value of data and providing protection for consumers and businesses.
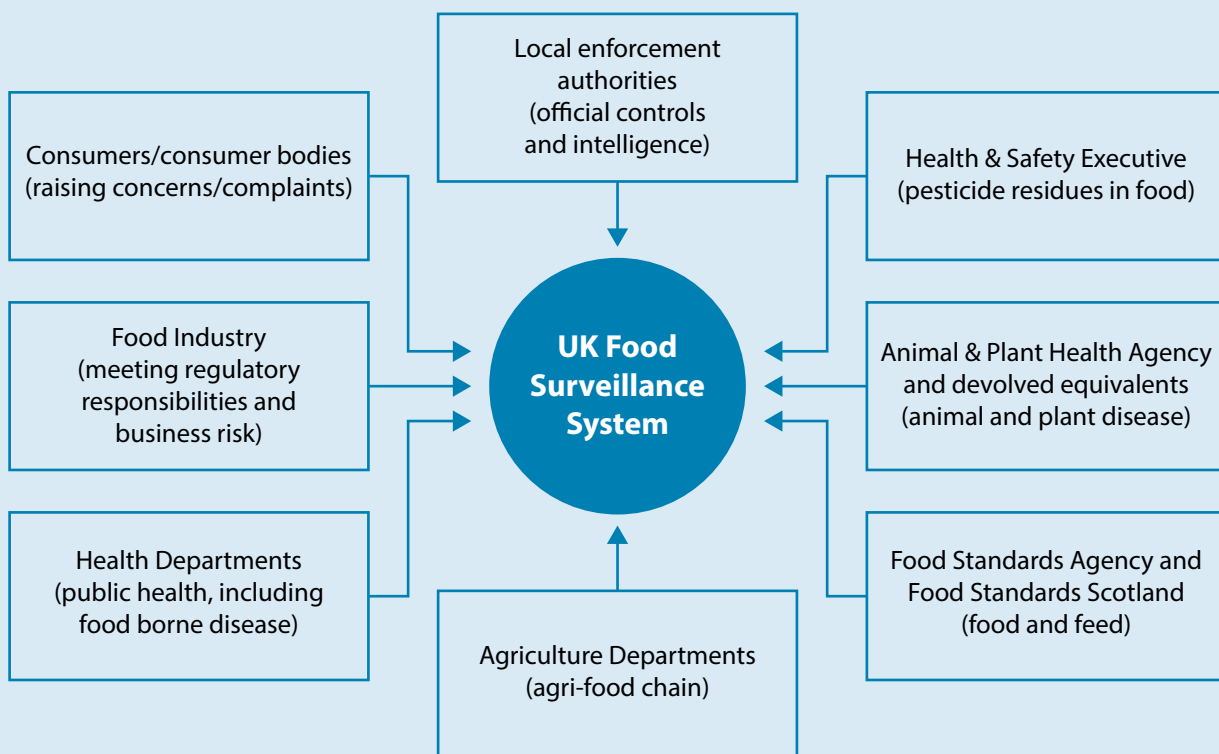
For more information, visit: **food.gov.uk**

f Let's keep connected at **food.gov.uk/facebook**

🐦 Join our conversation **@foodgov** using **#CSAreport**

▶ Watch us on **food.gov.uk/youtube**

Ⓟ Repin us at **food.gov.uk/pinterest**

V Revine us at **food.gov.uk/vine**