

Chief Scientific Adviser's Science Report

Issue eight: Data Standards



“Here at the Food Standards Agency (FSA) we aim to create a strong data culture, strengthening our position as a ‘data driven organisation’. On its own, data is of little value, and only becomes valuable when we can transform it into information and turn that information into action. However, to do this, it must be available quickly and be ready to use when an individual or team identify the need for it.

As an example, standardising allergy alerts has reduced the time taken to issue alerts to consumers through automation and the use of consistent, clear and unambiguous data. This means that those who need to know can find out in the quickest, most accessible means preferred.

In this report, we demonstrate how current and future work on data standards aims to make the data we create available, easy to access and supported by the necessary tools and expertise. Maximising the effect of our data is part of being an excellent, modern, accountable regulator, delivering food we can trust.”

Professor Guy Poppy,
FSA Chief Scientific Adviser

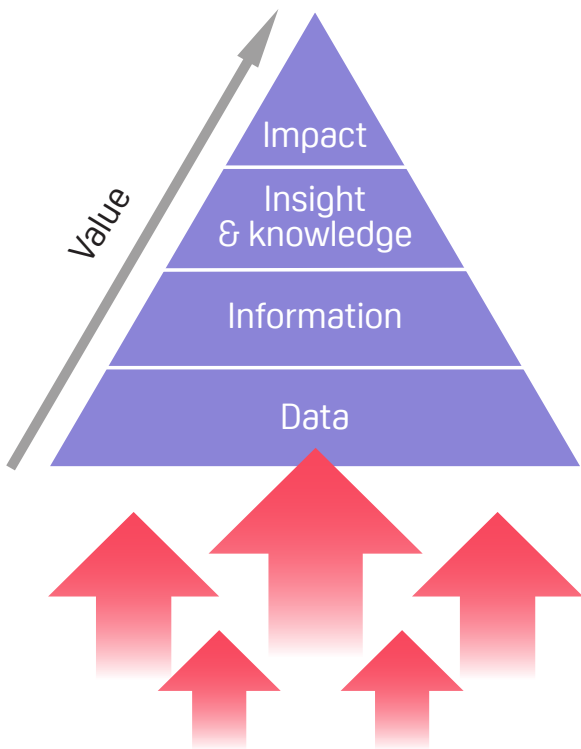


Our approach to data standards

Data collection and storage is growing at an astounding rate and is transforming our lives. More digital data has been created in the past few years than in the entire history of the human race¹. In 2012, the amount of digitally held data first exceeded a zettabyte (1 million terabytes) and in 2018 a white paper estimated that the total is now 33 times that².

To put this into context if you imagine a standard (325ml) cup of coffee as holding one gigabyte, a zettabyte would have the same volume as the Great Wall of China³. The total volume of data is only increasing and is doing so in a largely unstructured way. This data is a great potential asset. In the 2017 CSA Science Report on Data Science we talked about the data pyramid, which highlighted how data alone has little to no value.

The Data Pyramid



The data pyramid shows how we use data to create impact. Raw data, which can come from multiple sources lays the foundations. It is essential that the ‘right’ data is included. The value of data is most readily realised high in the pyramid through the conversion of knowledge to impact. By making our data open and making effective use of others’ data we strive to maximise value and impact.

It is essential that the ‘right’ data is included.

The report also discussed how most of a data scientist’s time can be spent obtaining, cleaning and reformatting data in preparation for analysis. Data standards are the most effective way to accelerate progress from data to information by reducing the amount of effort and elapse time needed to prepare data.

In this report we outline how important data standards are to the FSA, our approach to adoption and use, and provide real examples of where deployment of the data standards is making a material impact. Our approach to data standards can be divided into three categories:

Adoption

Where possible, we want to use recognised standards created and maintained by others including international standards bodies, fellow regulators, the food industry and others. Examples include using standardised data about climate and weather, standards for recognised identifiers like barcodes or food description ontologies such as [FoodEx2](#), or [international standards for recording times and dates](#) accurately.

- <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#576ea50860ba>
- <https://www.seagate.com/gb/en/our-story/data-age-2025/>
- <https://www.theguardian.com/technology/blog/2011/jun/29/zettabyte-data-internet-cisco>

Consensus

The value of standards increases as more people use them. For maximum impact, where possible, it is useful to reach consensus on standards with a broad group of users to develop them. It is crucial that we work with others to develop standards that meet needs without creating a burden. As standards are adopted the cost of being compliant reduces over time as those standards and their associated products become commodities. This could involve working with users to create and maintain

standards for a subset of the food industry, or agreeing and maintaining a controlled list of terms, managing and using that in data sets.

Ownership

In areas where the Agency performs a specific regulatory function, it is important that we set, manage and maintain clear and unambiguous data standards so we can effectively use the data we receive, publish and hold, and we enable others to do the same.

CASE STUDIES:

1. Adoption case study – Surveillance

FSA food surveillance is about building a picture of the food system, its risks and vulnerabilities, so that the FSA and others can manage consumer risks and support the delivery of safe and authentic food.

FSA food surveillance is about building a picture of the food system.

We are building a set of services, using open data and data science techniques, to identify emerging risks before they impact on public health. Data sources are identified, cleaned and curated so they conform to clearly defined standards, allowing them to be combined.

A recent project examined Aflatoxins – a toxin produced by certain fungi and found in some foods. We know that some foods

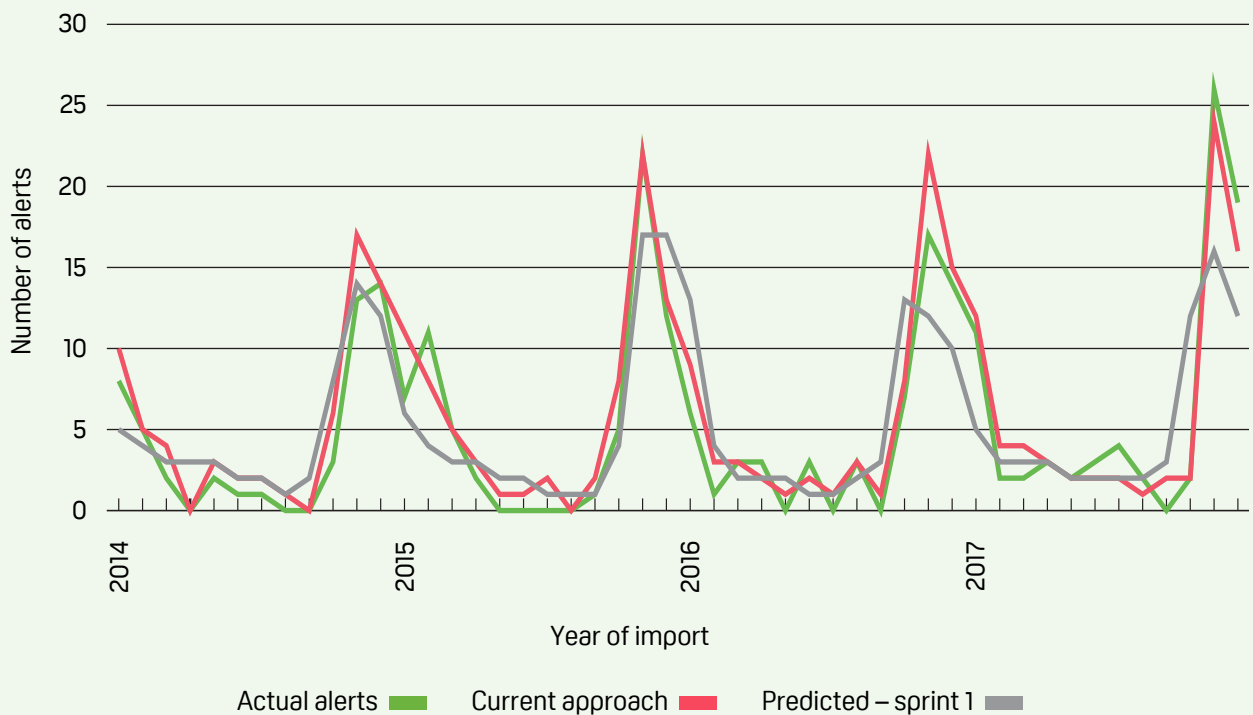
have higher levels of Aflatoxins at certain times of the year. Looking at climate data in specific geographical areas we have been able to build a predictive model that identifies increased Aflatoxin risk.

As climate data is standard for variables such as rainfall, humidity and temperature it is possible to identify other areas of the world which have similar climatic conditions.

This gives us important information about where other potential risks might arise, and we can direct our food sampling activity accordingly.

For this work we used open datasets from RASFF (Food and Feed Safety Alerts), HMRC trade data, and open climate data. We were able to use standard data across all these sources to link the information and create knowledge, providing a positive impact to our work in risk identification.

Hazard Occurrence profile: Aflatoxin hazard alerts



2. Consensus case study – Collection and Communication of Inspection Results

In 2016 the FSA embarked on an initiative to standardise the data collected as part of the Collection and Communication of Inspection Results (CCIR).

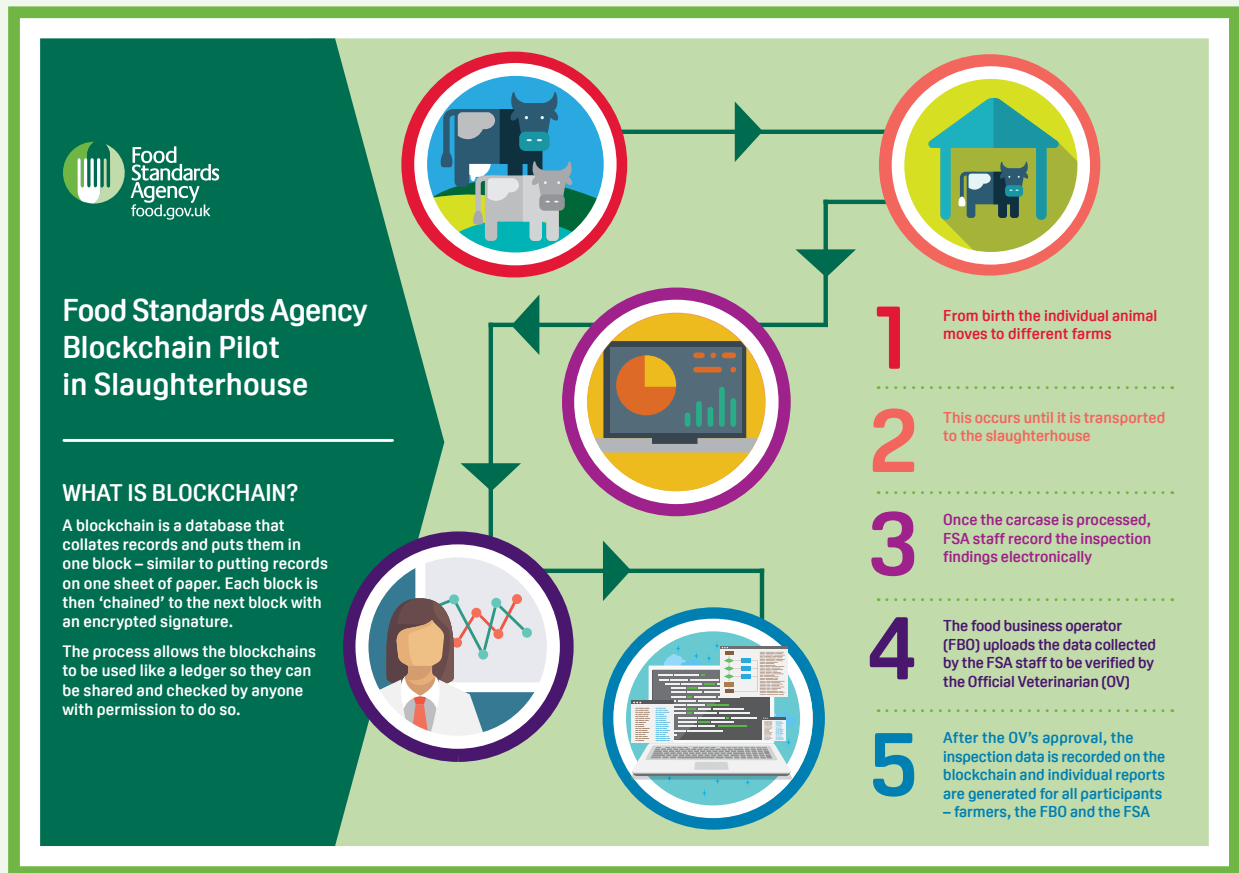
This was a detailed process of working with the meat industry to agree what animal conditions should be collected across all meat species slaughtered in the UK. As a consequence, the Agency has published a list of conditions collected at slaughter for each [species](#).

In 2018, through a pilot, we tested the use of a Distributed Ledger Technology (DLT) to make the CCIR process more efficient in a cattle abattoir. The approach used a permissioned ledger and modern

technology processes to take data from the abattoir Management Information system and write this to the blockchain. Whilst making use of emerging technology, this did not alter the statutory process of information collection and recording.



What is Blockchain?



Later in 2018 the pilot was rolled out to a further two abattoirs and the FSA is in the process of working with a further Food Business Operator to roll out in their plants, making us a UK Government leader in real world application of DLT.

FSA is a UK Government leader in real world application of DLT.

Concurrently, all involved are considering the implications of the approach and establishing an appropriate collaborative governance model to continue the work.

While the DLT capability is not deployed to its full potential in the pilot (processing at most, hundreds of data points a day in any plant compared with hundreds of thousands of times larger scale operations in Financial Technologies), the real value accrues because of the consistency of the data standards across plants and the standardised technology developed to write content to the blockchain.

Also critical in this use-case is the permissioned nature of the ledger which allows trusted individuals or organisations to access different versions of the information.

3. Ownership case study – Alerts

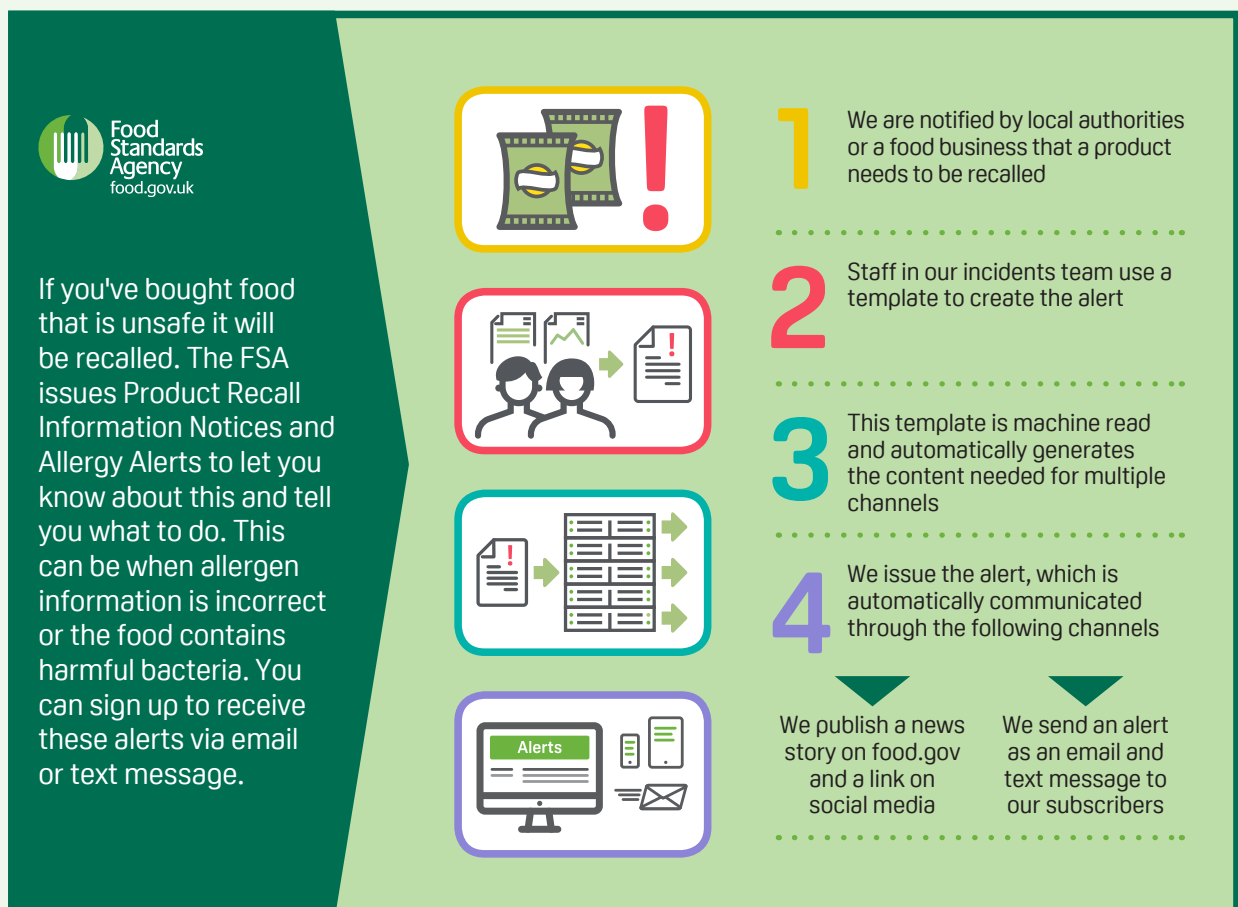
The FSA is responsible for issuing alerts of product recalls and withdrawals. This is some of the most impactful work that we do. Foods sometimes must be withdrawn or recalled if there is a risk to consumers. This could be because the allergy labelling is missing or incorrect, or if there is another kind of risk to the food, such as microbiological, chemical or physical contamination.

We have business processes for managing alerts, from notification of a potential incident to the publication of an alert. Over the years, some standardisation has developed within those processes, templating the alert and

providing structured tables for describing products affected. We wanted to improve the issuing of alerts by automating as much of the process as possible. Time is of the essence here, to warn consumers as quickly as possible of what can be risk to life.

Data standards have been applied to make alerts more consistent, which in turn reduces the time taken to issue the alert. This means that the data is easier for developers to work with in their services. In line with the government standard approach, we use this data to power our own alerts webservice. This is achieved by developing detailed data standards for the different types of alerts we issue, making them machine readable, and using

Food and allergy alerts



those standards to develop an application programming interface (API) to make the data machine accessible.

One element of the process that has benefitted from data standards is how we refer to an individual alert.

One element of the process that has benefitted from data standards is how we refer to an individual alert. Previously each alert could have had a different format of reference number, or worse, a reference for one kind of alert could have been repeated in another which could cause confusion for consumers. We now use a simple pattern for reference numbers which makes them machine readable and understandable to users at a glance. Users can now tell the type of alert and its recency from the structure of the alert reference alone.

Creating functioning data standards often starts with work to create repeatable, reusable patterns for data fields like reference numbers or code lists for items that require repeatable and consistent ways to manage terms.

This work brings the following benefits:

- we issue more consistent alerts, which are faster to produce and publish
- staff with less subject matter expertise can create and manage alerts, because the data structures help to guide them through the task
- fewer transcription errors occur in the data fields as entries are pre-set
- third parties can access the data at the same time we publish it, extending the reach of our alerts through other services and apps.

Registers

Registers are canonical sources of data that can be trusted to be accurate, up to date and managed by the people with the best understanding of the subject matter.

They are a key part of the Government's approach to data as infrastructure. Much like a road can be relied upon to be functioning and get you from point A to point B because of the commitment of the owners to maintain and update that infrastructure, registers do the same thing for data.

We have worked closely with the Government Digital Service to create a [register of allergens](#).



The register provides consistent definitions for each of the 14 allergens in food products, that must be declared by food businesses, so that services and users can share the same definitions.

The register can also be updated with new allergens should the law change in the future. Creating consistently named, official sources of reference data is one of the best

ways to support data standards and make data sets interoperable.

The standards are obligatory for businesses. They recognise the ease and value the clarity provides when dealing with the regulator or indeed any of their supply chain, when the same standards are being used. The FSA believes this standard setting is a key role of the regulator.

A Unified View

One of the problems we encounter often is that useful data – whether that’s internal or coming from our partners – is less structured than we would like, or in a different format to the one we prefer to use.

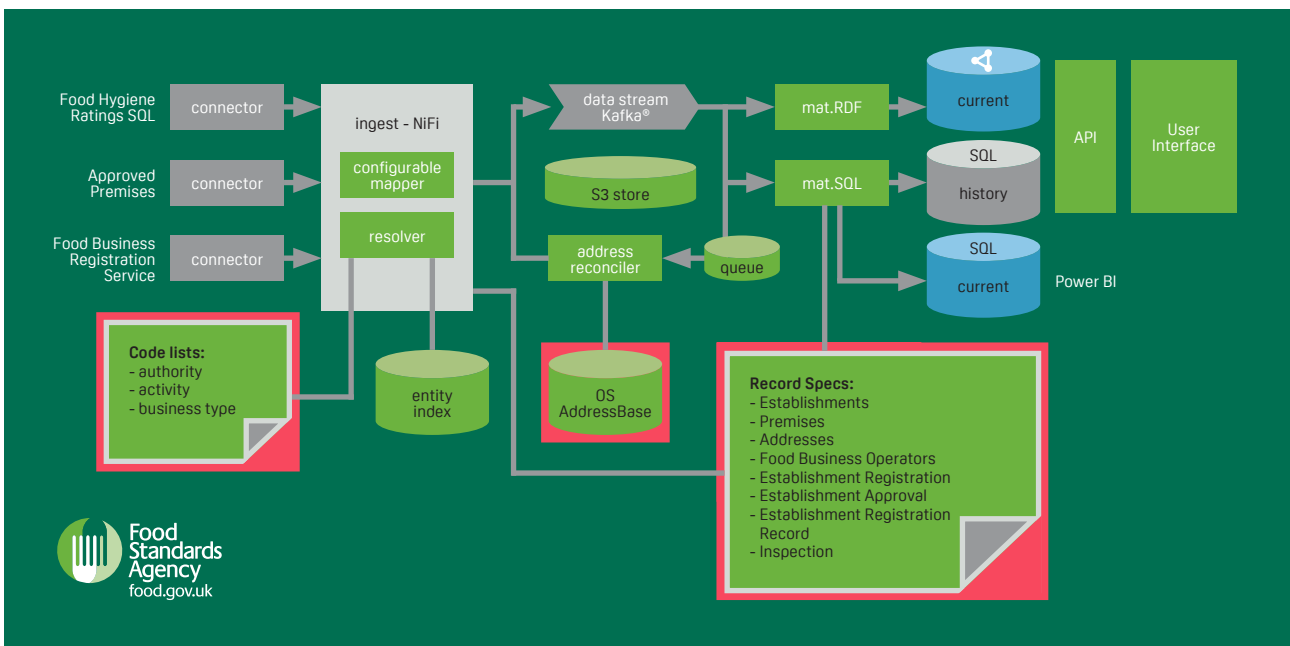
This is generally the case where the data has been collected or generated by people who need it for some function of their role, so they are collecting what they need in the way they need it to get their job done.

In an ideal world, everyone in our data ecosystem would collect data in exactly the

way we like it and provide it in our preferred formats. In this scenario though, our needs add no value to anyone else’s tasks and we need to work with what we get.

In an ideal world, everyone in our data ecosystem would collect data in exactly the way we like it.

The creation of the Unified View is work in progress. It will deliver services that bring together all the datasets we consider crucial in our role as regulator and transform those data



into the format and structure we want. This isn't a new process in itself, as teams of analysts have been manually processing spreadsheets and files for years to provide regular reports across the Agency. What's new is:

- our use of more innovative tools and technologies to access, validate and structure the data as we ingest it
- a central point to access and visualise the data on demand, with appropriate levels of permissions
- automation of the processes we use to fetch the data and make it visible.

Using these techniques, we can reduce our internal storage and processing burden while making sure we always have the most current and complete version of others' data.

We can meet the needs of internal data users across a range of business areas, whether that's sophisticated dashboards, simple lists, detailed statistical analyses or ad hoc queries. We are also using the improved data to feedback to the originators so they can make improvements at source, creating a positive feedback loop.

The value of data standards

Data standards do not mean we can turn data into action immediately, but they help us expedite the process. Quality data is data that is consistent and fit for purpose. Not all data is useful for all users, but clear standards enable users to make better judgments about whether that data is suitable for their needs.

There is a cost to establishing data standards and embedding them in the processes and culture of an organisation, but the additional value they can bring to improving processes in the short term, and insight in the long term, is worth the investment. Datasets are no longer just a number of assets (or liabilities) but together form part of the infrastructure of a decision-making organisation.

Data as infrastructure is a critical utility in a world where the nature, volume and accessibility of data is exponentially increasing. We can further capitalise on opportunities in artificial intelligence (AI) and machine learning (ML) within the food system. These technologies are already making an impact on our daily lives in other areas; from analysing and predicting our entertainment choices and shopping habits to developing autonomous

vehicles. AI and ML processes are data hungry, meaning that investment in data as infrastructure gains importance because the time that teams are spending sourcing, cleaning and manipulating data is wasted time.

Data standards are not static, they evolve over time as user needs change. For this reason, it is key that when we develop data standards we do so with extensibility in mind and continue to scan for relevant and useful standards that we can adopt before (re)creating a standard.



Also of interest: The Science Council

Science Council Working Group on Digital Technology and Data Usage

The FSA's [Science Council](#) was established in 2017 to provide independent strategic advice and challenge on how the FSA uses science.

The Council currently has a Working Group addressing the following 4 key questions on data usage and digital technology:

1. Over the next 2-5 years, what are likely to be the emerging data tools, techniques and technologies which could have the greatest impact on the FSA's mission, including for its Regulating Our Future programme, and what value could these add?
2. Where and how could the FSA benefit from modifying its data collection processes? What are the expected benefits of any changes?
3. What are key implications for the FSA of advancements in open data, data sharing and how should the Agency go about leveraging them (including private / industry data)?
4. How can the FSA ensure that it adopts the right controls and governance

around data? This will be informed by a review of data governance and legal & regulatory frameworks for the use of data in decision making (including pre-disclosure and pre-disposition in relation to machine learning and artificial intelligence) and, associated with this, the ethical use of data.

The Working Group is led by Professor Patrick Wolfe, who is a member of the Science Council and the Frederick L. Hovde Dean of Science at Purdue University. At a recent meeting of the Science Council, Professor Wolfe explained that:

“Data is a crucial topic for the FSA and really underpins the important work that the Agency does. Understanding how to capitalise on the data the FSA holds will help the organisation to realise the art of what is possible at the leading edge of technology, and to produce a strategy that will deliver the benefits of data technologies directly to the consumer.”



Professor Patrick Wolfe

Glossary

SQL	Structured Query Language
OS	Operating System
S3	Simple Storage Service
RDF	Resource Description Framework
NiFi	Apache NiFi

Acknowledgements

With thanks for the contributions from Julie Pierce, Siân Thomas, Adam Locker, Lucy Knight, Jesús Alvarez-Pinera and Suzanne Fox.

For more information, visit: [food.gov.uk](https://www.food.gov.uk)

 Let's keep connected at [food.gov.uk/facebook](https://www.food.gov.uk/facebook)

 Join our conversation [@foodgov](https://twitter.com/foodgov) using [#CSAreport](https://twitter.com/foodgov)

 Watch us on [food.gov.uk/youtube](https://www.food.gov.uk/youtube)

 Repin us at [food.gov.uk/pinterest](https://www.food.gov.uk/pinterest)