
Food Standards Agency
food.gov.uk
Cognizant
Predicting Food and Feed Risks with the use of Data Science
20th March 2019

# We are:

## Jesus Alvarez-Pinera

Head of Strategic Surveillance at FSA

Since mid-October 2017, I have been the FSA lead on the Agency's new Strategic Surveillance.

I lead the development of an operating process to collect input on risk and vulnerability signals across the organisation and externally. Using available data/intelligence we aim to identify risks at an early predictive stage to drive initiatives, policies and interventions to mitigate those risks before they become incidents.

Before the FSA, I worked for several years in academia, as a Cell Biology professor, where my area of expertise was the early stages of the formation of the skeleton. During this period, I published over 30 articles in international scientific journals.

## Nitin Gangwani

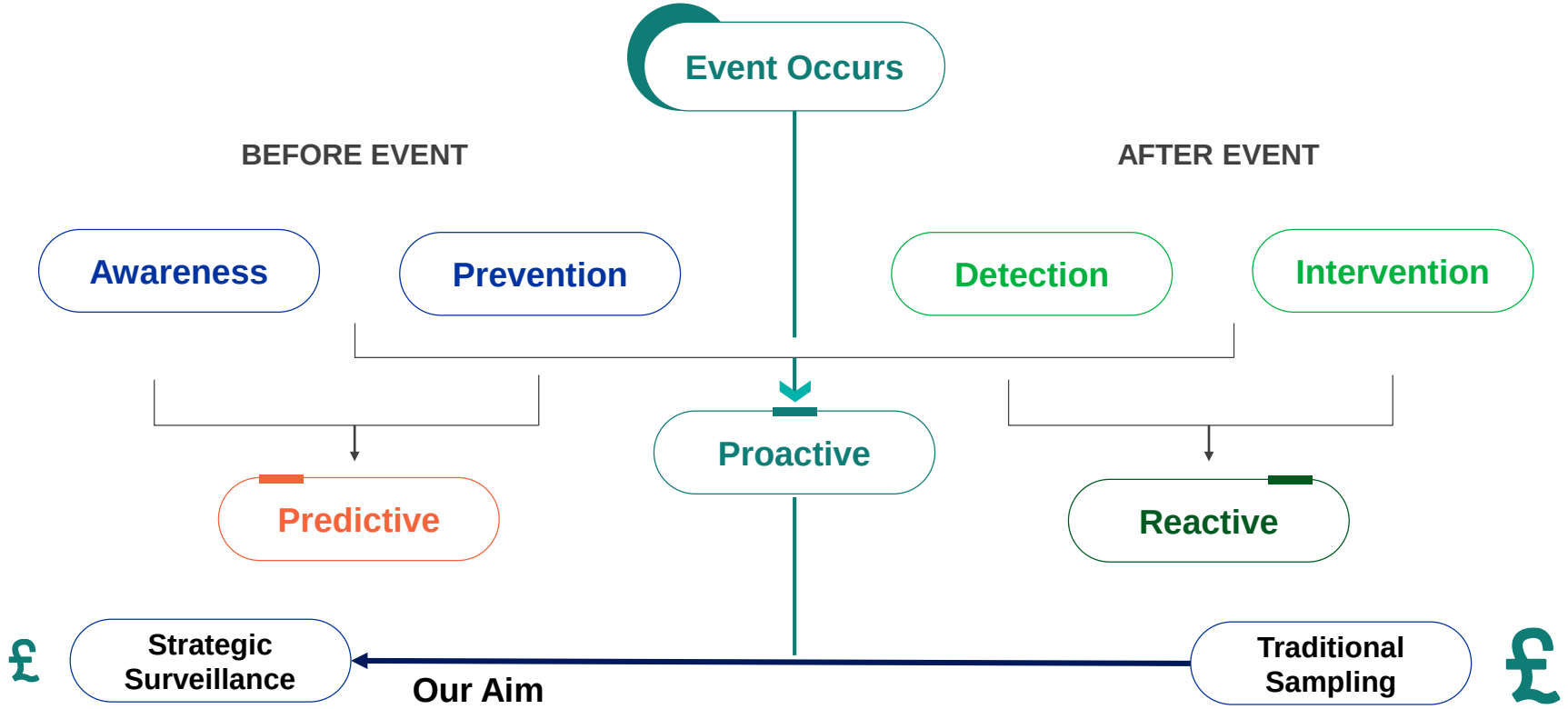
Senior Manager and Project Lead at Cognizant

I am the Cognizant project lead working on the FSA strategic surveillance initiative since June 2018.

I have been with Cognizant for over 6 years delivering analytical solutions to drive business transformation through data science. At the FSA, I lead the team which develops solutions to deliver actionable data-driven insights to meet our stakeholders' requirements in line with the FSA's strategic surveillance vision.

Previously, I have worked on and managed multiple engagements in various locations such as Australia, Japan and India, delivering analytical consulting and technical solutions in other industries such as manufacturing, insurance, life sciences, healthcare, etc.

# From traditional sampling to Strategic Surveillance

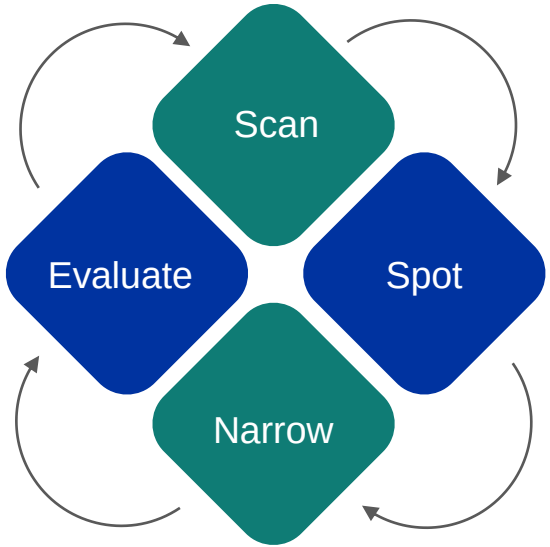


# Strategic Surveillance

**Our aim** is to make better use of (open) data to identify emerging risks before they become a risk to public health and to take data-driven actions, e.g. provide insights that can guide better use of sampling resources.

**Our operating model** is the WHO's 'scan, spot, narrow, evaluate', improved by insights from RAND and learnings from previous use cases.

We follow a mature agile way of working that is centered around specific 'use cases' (projects).



# Surveillance vision built on data

To protect the consumer, and ensure that food is safe and authentic, we need to analyse data to understand where issues may arise **before** they are issues

**Protecting the consumer**

Food is safe to eat

Food is what is says it is

Identify emerging risks

Spot the anomalies

Take data-driven action

Assess historic incidents' data

Identify root causes of incidents

Share insights with partners

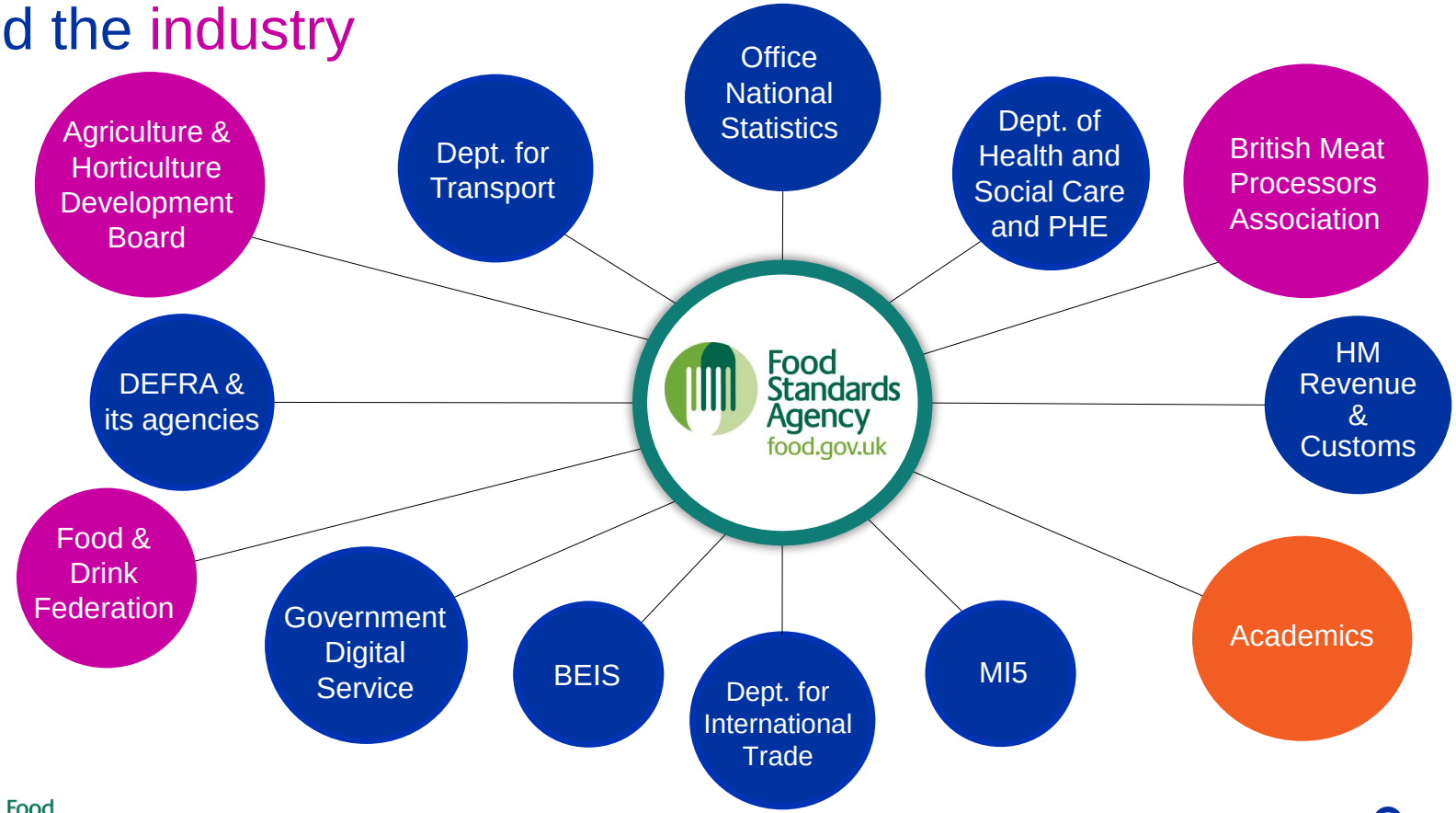
Understand impact of potential actions

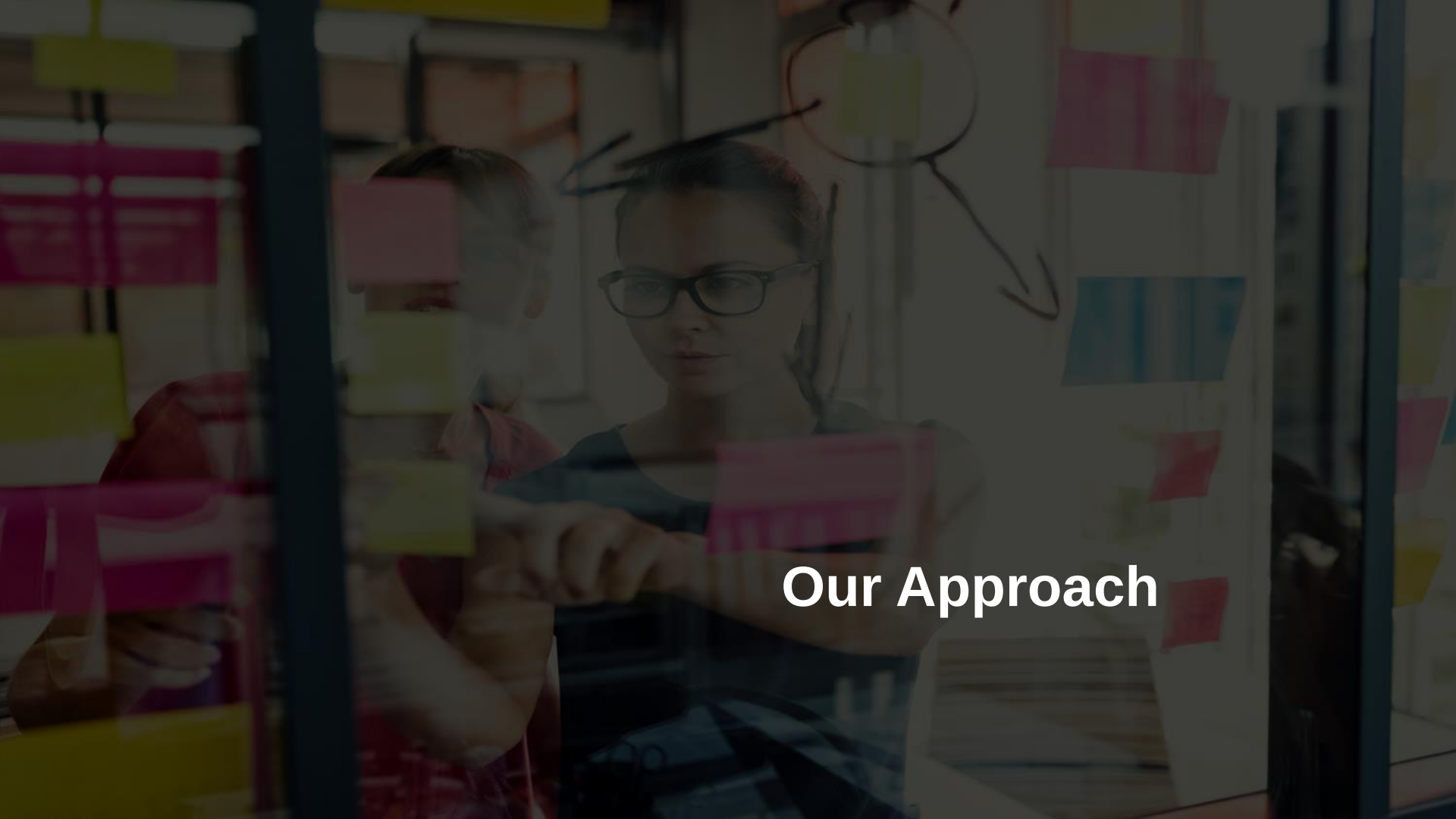
# Where do we want to be?

**We aim to build a Surveillance Capability (core service) that:**

- Supports the wider ambition that food is safe to eat and is what it says on the tin
- Helps us to **understand risks (safety / authenticity / assurance)** and identify both **gaps and risks** that are changing or not being managed, followed by a plan of action
- **Drives decision making** and prioritisation across all parts of the FSA and beyond
- **Uses evidence-based analytics** to deliver the appropriate level of confidence / certainty to drive decision-making

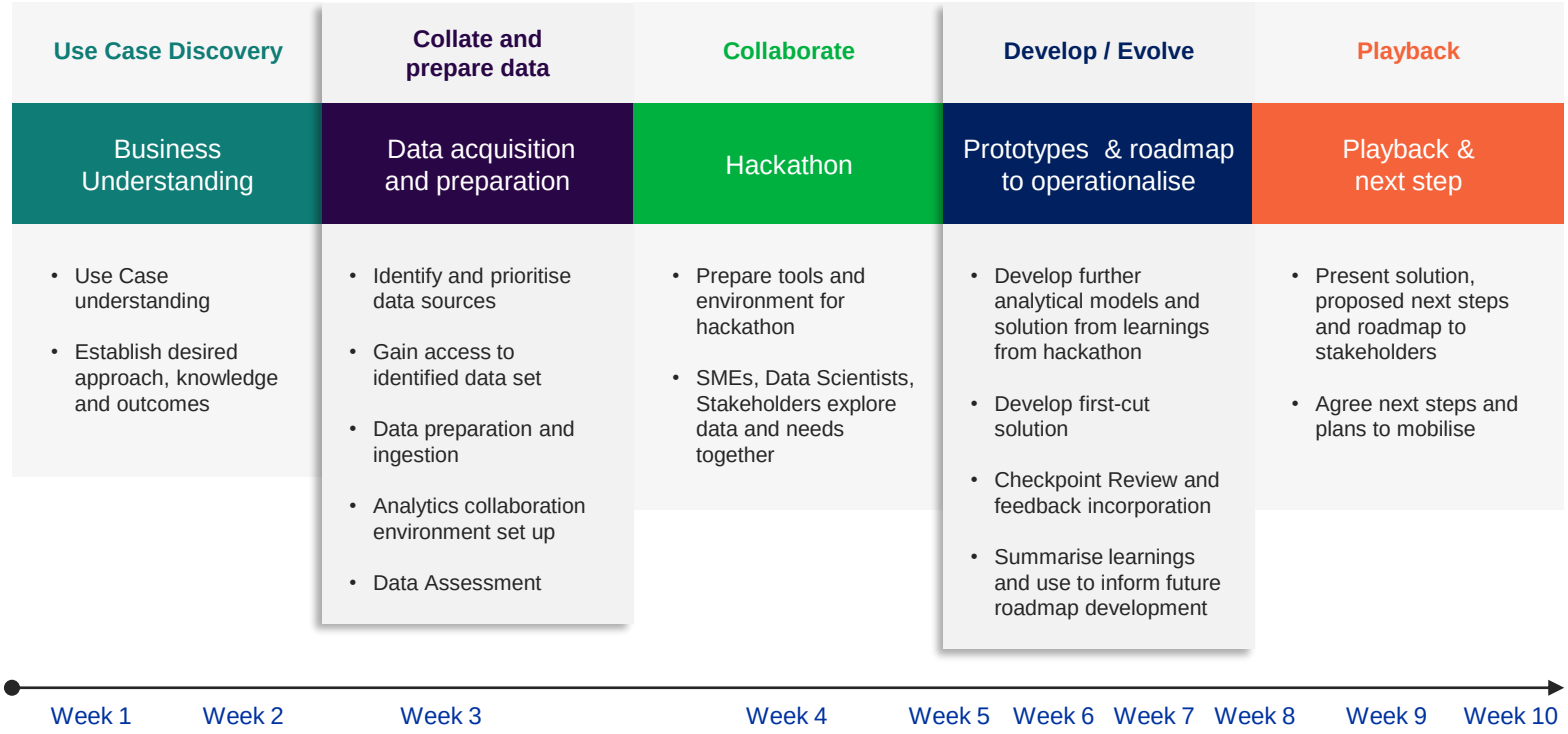
# Working together with government departments, academics and the industry





# Our Approach

# We work in well-defined agile sprints



# With the key to quickly generating insight

1. One of the key accelerators for strategic surveillance is the organisation of a **#hackathon** within each sprint, where subject matter experts & data scientists work together, and use data to create actionable insights

2. In order to be successful, we have realised a number of skills and experts across the business need to come together

3. Through collaboration, identifying valuable data sets becomes easier

4. Collaborating with business stakeholders throughout the development and prototyping phase is the key to creating a successful end product which could be utilised by them

5. Working on prototypes also highlights what went well and what didn't go so well, with this insight we can begin to map out a repeatable process, applying and testing the process with each cycle to refine and become increasingly effective and efficient.

# With a flexible approach


- we work in agile sprints, through prototypes
- we are flexible, responsive, speedy, cost effective
- with open data, open source algorithms
- with business, academia and government
- whether here and now, or finding the unknown unknowns
- we are starting to predict risks, not just find them on the doorstep
- we use technologies and data science through machine learning, predictive modeling etc.

# Examples of our Work

# Most talked about allergies on social media

## Problem Statement

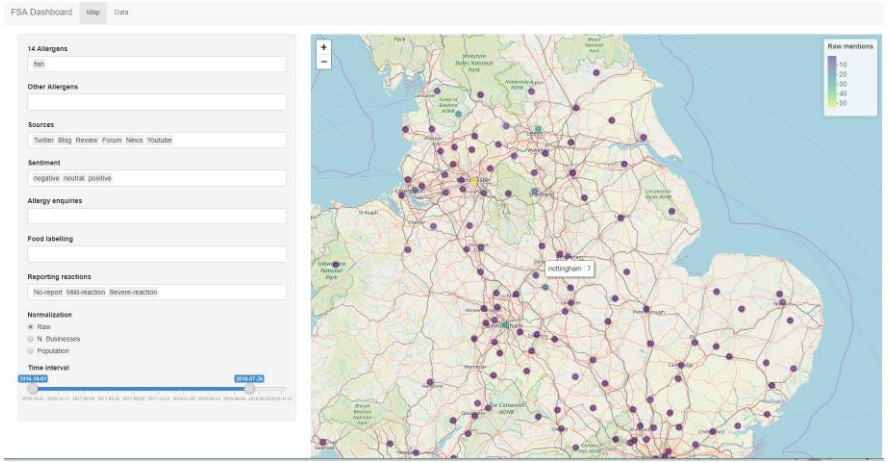
What can we understand about allergies, using social media?

## Solution Highlights

- Focussed on two questions:
  - Which allergens are people talking about the most?
  - Which allergy related issues are people talking about, and is there any difference across local authorities?
- Analysed 2 year's of social media data

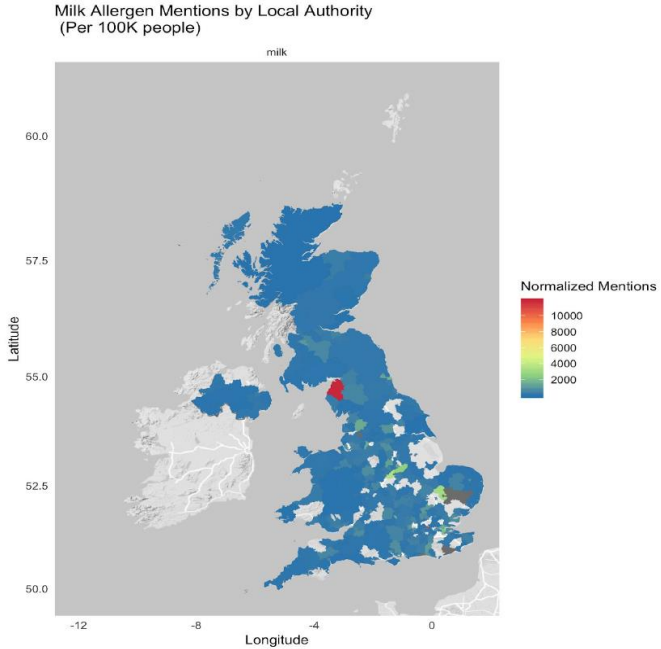
## Key Findings and Way Forward

- Identified most discussed allergens – Coconut in top 10
- Looked at issues by local authority
- Created dashboard to interact with data

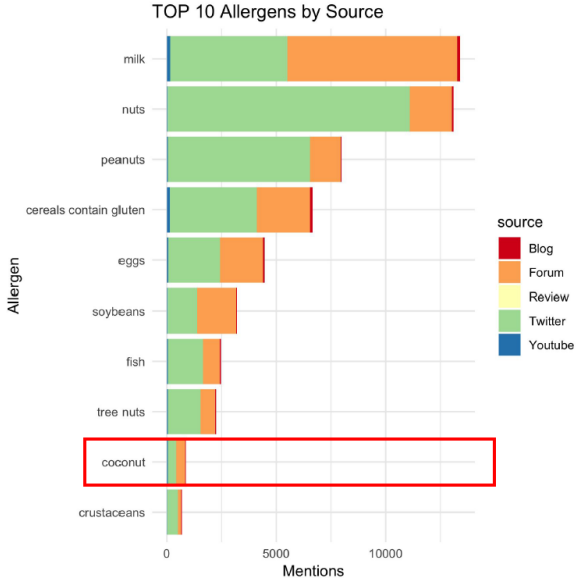


# Most talked about allergies on social media

How do allergen mentions differ among Local Authorities?




Are other allergens outside the list of 14 being talked about?



# Identifying imported food from EU and non – EU countries which could be risky but are not under official controls

## Business Objective

Identify and prioritise list of risky products originating from EU and non-EU countries which are not a part of the current regulations but are risky.

## Solution Highlights

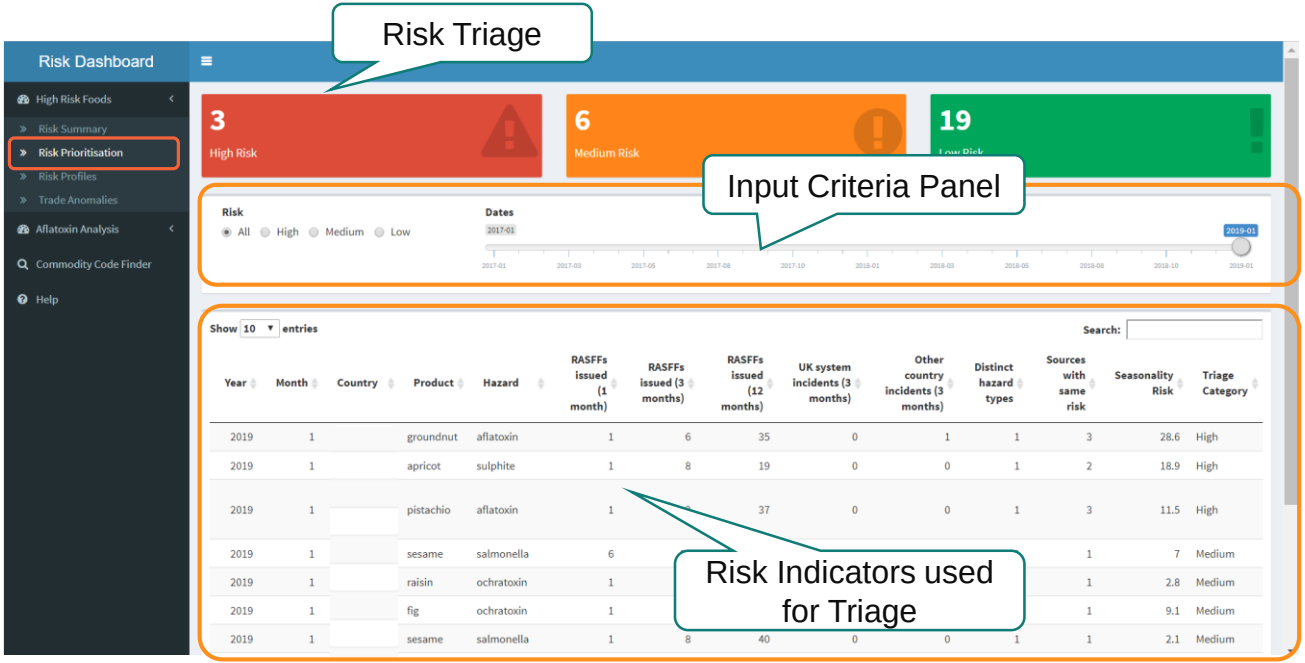
Use machine learning models to determine the list of risky commodities at commodity-country-hazard from various data systems:

- EU Rejections (RASFF)
- FSA Data
- Local Authority Data
- Lab Data
- TRACES
- Other public data
- Other Countries – Japan, US, Canada, Australia etc.

Develop various risk indicators to understand the likelihood of risk to prioritise risk. Example:

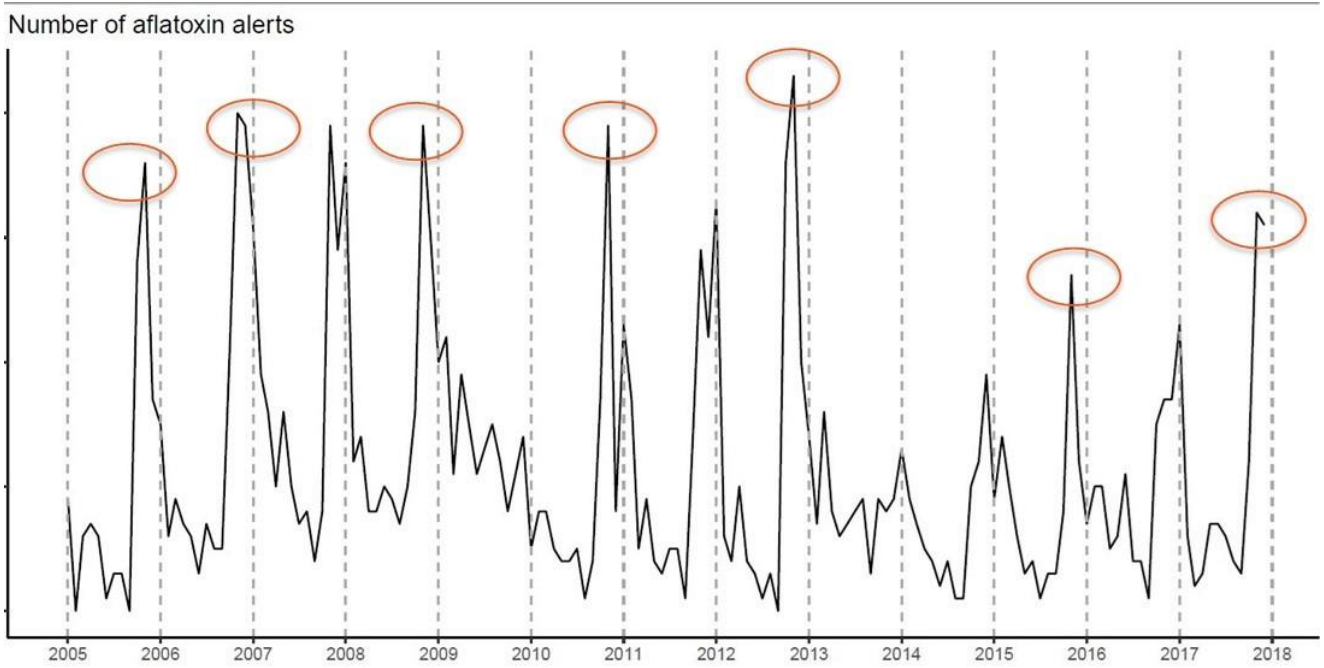
- # Historical alerts (in past 1, 3, 6 months etc.)
- Trade Volume / Exposure
- Rejections from official sources in other countries
- Price / Trade anomalies
- Seasonality
- Etc.

# Using ML algorithms, we triage risks into buckets using indicators and highlight critical issues



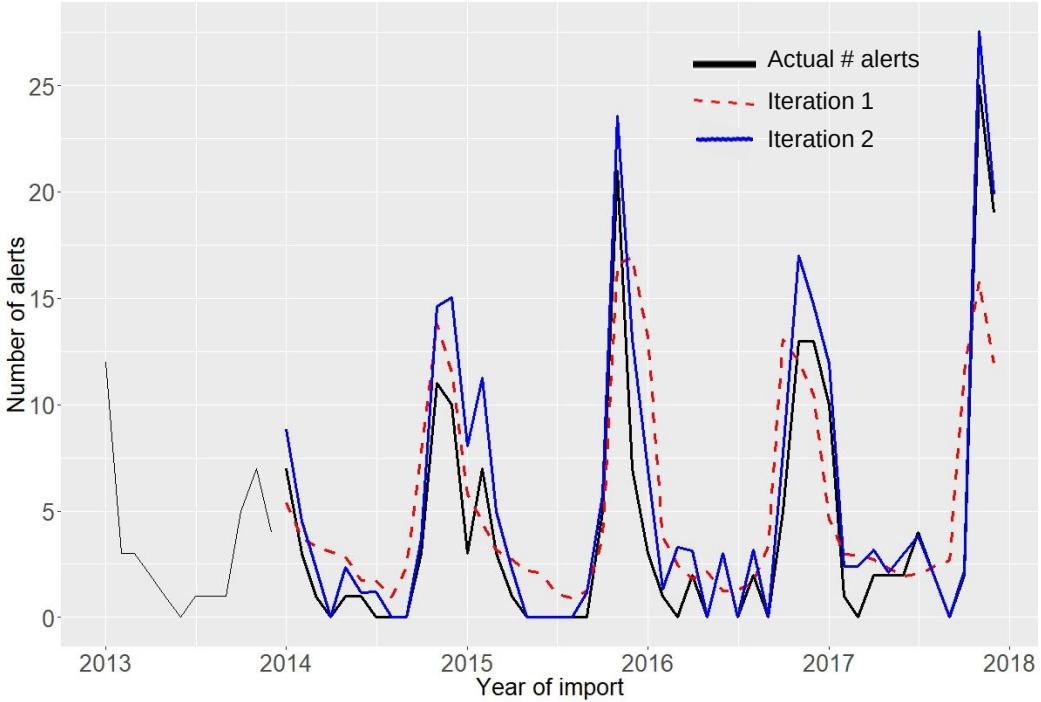

The dashboard provides the end users with a triage for various food and feed risks seen across various data sources. By developing certain indicators, and using ML algorithms, final triage category is determined

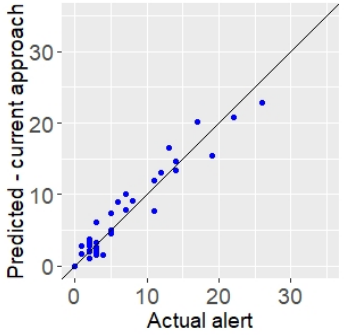
# Understanding the seasonality of Aflatoxins risk in commodities through a predictive model




- In this data, months – October and November have higher number of incidents (due to Aflatoxins)
- We wanted to understand the root cause of the seasonality patterns and build a predictive model which could help us be aware of the risk in advance

# We built an analytics model which predicts Aflatoxin risk based on historical data and climate conditions




Adjusting the alerts per million Tons imports removes the volume effect and remainder can be inferred as the true effect of weather in causing Aflatoxins

Variables such as temperature, rainfall, harvest month, storage months etc. were utilised to build the model based on the latitude and longitude of the area where the commodity was most likely grown

# Moving from known knowns to unknown unknowns for the Aflatoxin model

## First Iteration

**Minimal viable product (MVP)**

The absolute minimum we develop to create insight – 'Predict Known Risk' by utilising climate data

Aflatoxin risk prediction model using RASFF for one specific commodity (regulated – known risk)

**Known knowns**

## Second Iteration

**Generalised model**

**Second Iteration:** We generalise model to ensure relevance to wider issue / risk

Model using RASFF to predict aflatoxin alerts for commodities with Aflatoxin risk in regulations 669, 884 etc.

**Known/unknown unknowns**

## Third (Recent) Iteration

**Business as usual (BAU)**

**Third Iteration:** Extend the model to predict risk from unregulated countries (unknown risk)

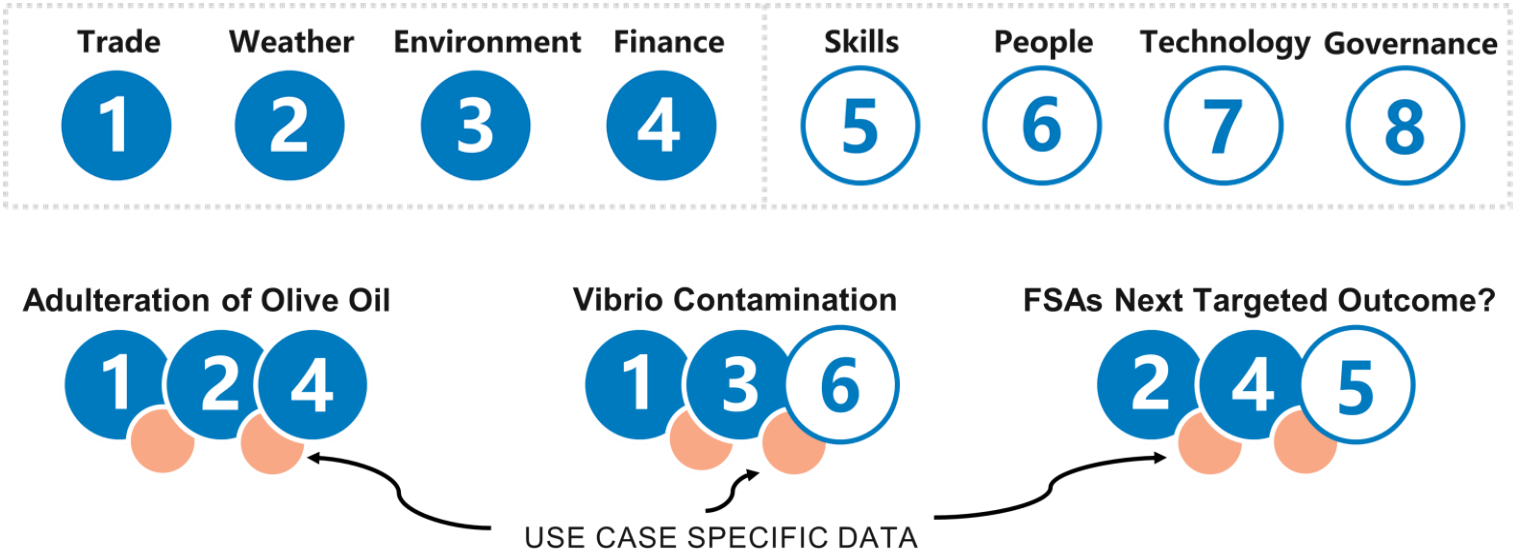
Generalised Model to predict Aflatoxin risk for food imported independent of country or commodity

**Unknown unknowns**

# What: Strategic Surveillance is transferable

- Data algorithms and machine learning models can be re-used, either in part or whole – transferable models and common data sets will over time result in a ‘toolbox’ and act as accelerators.

- We can already identify some datasets expected to be common  across many use-cases, such as Trade, Climate and Environment.



# The Benefits

# The benefits provided by Strategic Surveillance

Predictive risk insight using robust data science techniques - scan range of data sources, spot signals, narrow down and delivery using evidence based analytics

Use of innovative data science driven tools to empower and inform FSA customers of existing and/or emerging food risks

Empower FSA customers with key insights into their requested business problems

Support targeted interventions by enabling effective sampling with reduced overheads via analysis of specific threats

Provide proactive guidance, direction and specialist knowledge across FSA customers via their team of specialist expertise

Enable the FSA to respond rapidly, cost effectively and incrementally to emerging trends

Prevent food from entering the food chain or early mitigation by identifying a number of emerging food safety and authenticity threats

Focus on emerging threats to food safety by sharing insights and alerts across teams within the FSA

Food Standards Agency
food.gov.uk

Cognizant

# Questions for us?

# Thank you