



Funded by
the European Union

LIVING LAB #1

Identification and monitoring of food safety risks

2nd Annual Workshop (online)

17 June 2024

Chatham Rule

Participation in this meeting implies adherence to the Chatham House Rule:

“When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed.”

Registration to the HOLiFOOD stakeholders participation page



<https://bit.ly/HOLIFOOD>



Welcome

HOLiFOOD Living Lab #1 workshop,
June 17th, 2024

Schedule

- ▶ Introduction
 - ▶ Welcome and introduction
 - ▶ Overview of AI in emerging risk identification
 - ▶ Transfer to break-out sessions
- ▶ Break-out session 1
 - ▶ 2.1 Prediction models
 - ▶ 2.2 Text-mining models
 - ▶ Raw signal miner, Topic modelling
- ▶ Break-out session 2
- ▶ Final question and wrap-up



The HOLiFOOD project

Welcome and introduction to living lab #1

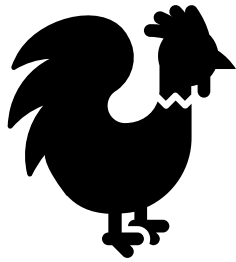
Overall strategic objective

“improve the integrated food safety risk analysis framework in Europe”

- ▶ “meet future challenges arising from Green Deal policy-driven transitions
 - ▶ in particular in relation to climate-driven changes,
- ▶ contribute to the UN Sustainable Development Goals 2, 8, 9, 12, & 15
- ▶ support the realization of a truly **safe and sustainable** food production”



Three selected supply chains



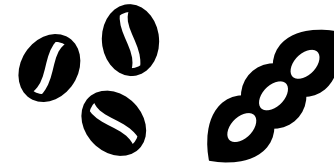
Poultry

[chicken]



Cereals

[maize]



Legumes

[lentils]

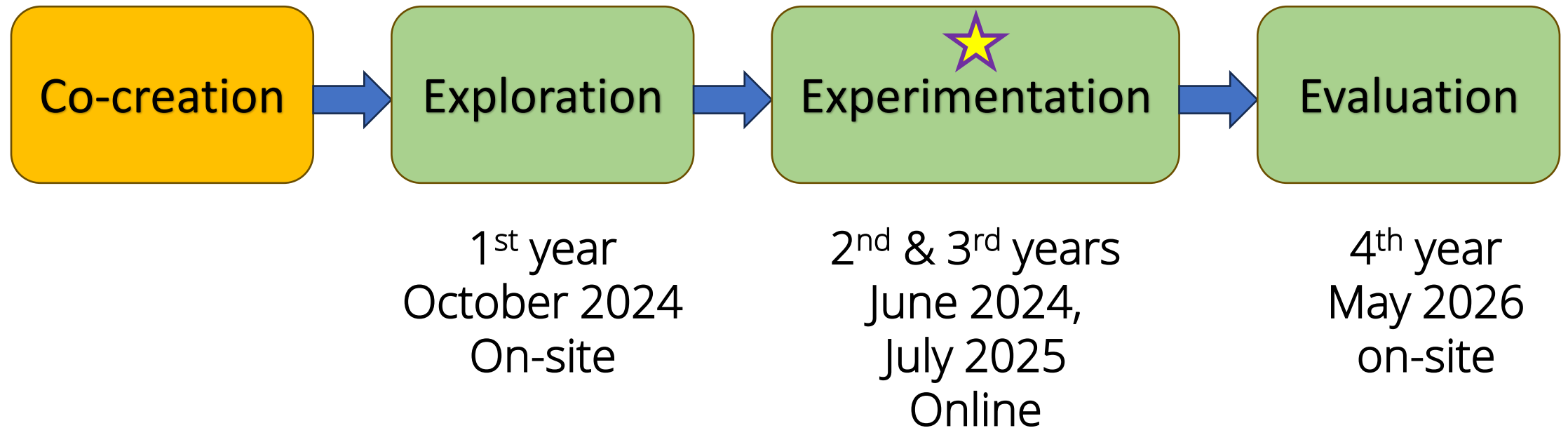
Living labs

- ▶ Bridge the gap between research and practice
- ▶ Multi Actor Approach (MAA)
- ▶ Fill the gap between development, validation, and implementation
- ▶ 3 Living Labs within HOLiFOOD:
 1. Identification and monitoring of food safety risks (WP1)
 2. Holistic risk assessment and acceptance (WP3)
 3. Platform co-design (WP6)

Living lab #1

- ▶ Supports Work Package 1, “Big Data technologies and Artificial Intelligence (AI) for food safety detection and prevention”
 - ▶ Development of dashboard integrating different tools
- ▶ Identifies the **needs of stakeholders** as related to **models and tools** for emerging risk identification (ERI).

Living lab #1 process



Exploration workshop (Synergy Days, Thessaloniki, 2023)





Funded by
the European Union

Emerging risk identification

Concepts, definitions, methodology

Ákos Józwiak, Zsuzsa Farkas

University of Veterinary Medicine Budapest

HOLIFOOD Living Lab 1: Identification and monitoring of food safety risks

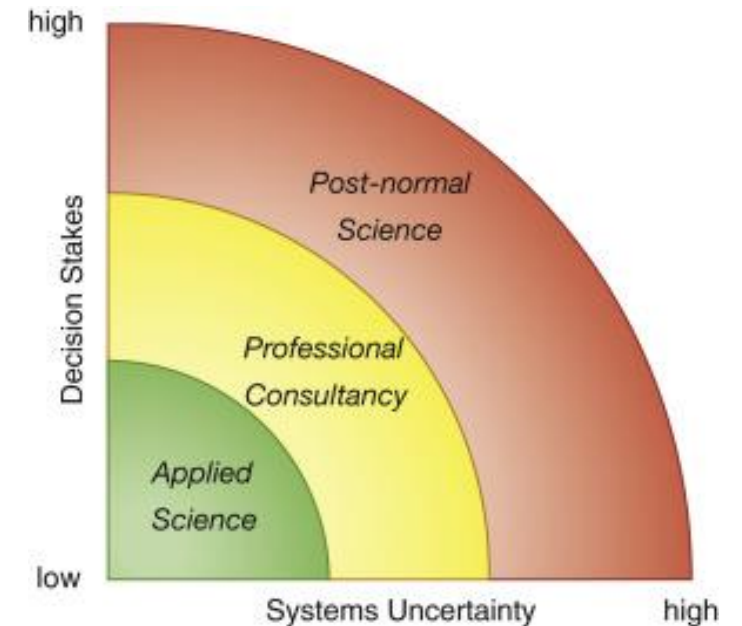
17 June 2024



Introduction

Risk based approach

- ▶ Mitigating food systems risks needs a profound knowledge on the prevalence and severity of risks
- ▶ Decision making problems in post-normal science:
 - ▶ based on uncertain facts
 - ▶ disputes over costs, ethics and values
 - ▶ urgent decisions needed
 - ▶ that may have far-reaching consequences
- ▶ Continuous evaluation of risks and continuous knowledge generation → **complex process**



Source: Silvio Funtowicz, Jerry Ravetz: Chapter 2 - Post-Normal Science: How Does It Resonate With the World of Today? In: Science for Policy Handbook, Elsevier, 2020. <https://doi.org/10.1016/B978-0-12-822596-7.00002-4>.

How to be prepared for the future?

POLITICAL ENVIRONMENT

FOOD SYSTEM STRUCTURE

CONSUMER ATTITUDE

GLOBAL TRADE

TECHNOLOGICAL DEVELOPMENT

CLIMATE CHANGE

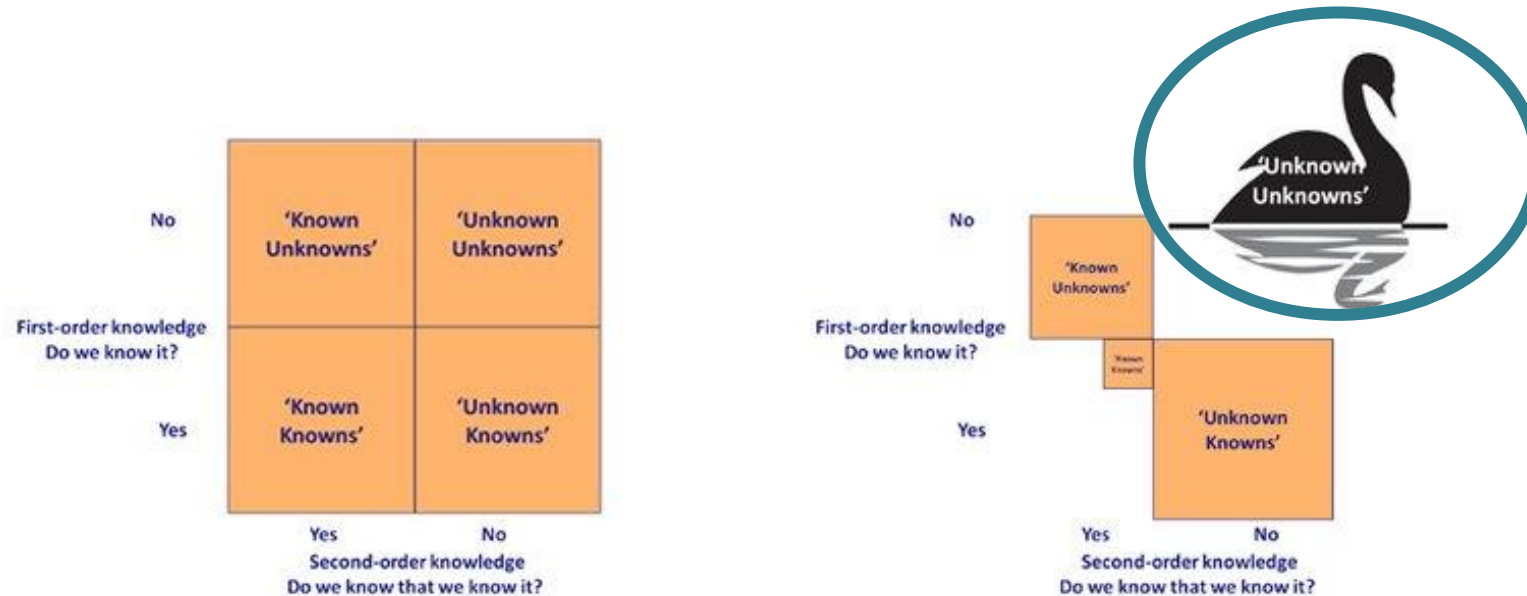
ECONOMIC GROWTH

DEMOGRAPHY AND
SOCIAL COHESION

DEPLETION OF
NATURAL RESOURCES

How to anticipate new issues?

- ▶ Finding the “next big food systems issue”
- ▶ Systematic analysis of short, medium and long timescale data and information



Horizon scanning universe



EARLY WARNING

SHORT TERM

Rapid alert systems

Immediate action required



EMERGING RISK IDENTIFICATION

MEDIUM TERM

Screening systems

Increases preparedness

Initiates risk assessment

FORESIGHT

LONG TERM

Driver and scenario analysis

Affects strategic actions

Early warning systems



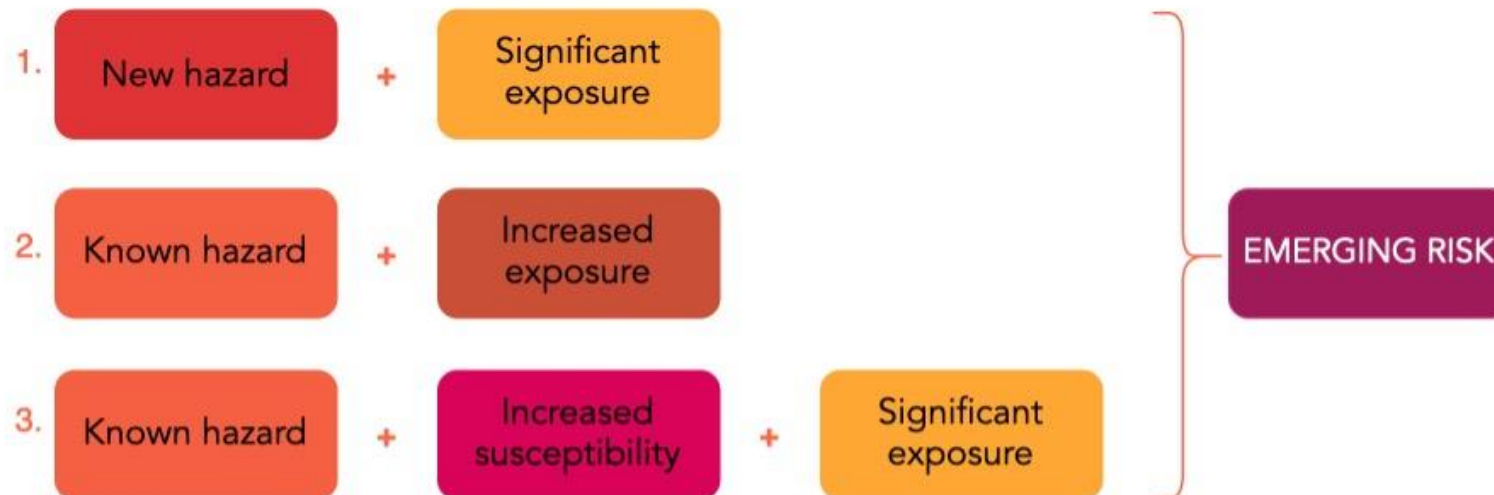
- ▶ Ongoing outbreaks/incidents somewhere else
- ▶ Share information → Take immediate action
- ▶ Rapid, structured information flow is required



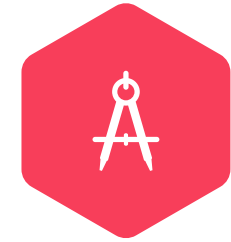
Emerging risk identification



- ▶ 'An emerging risk to human, animal and/or plant health is understood as a risk resulting from a newly identified hazard to which a significant exposure may occur or from an unexpected new or increased significant exposure and/or susceptibility to a known hazard' (EFSA)

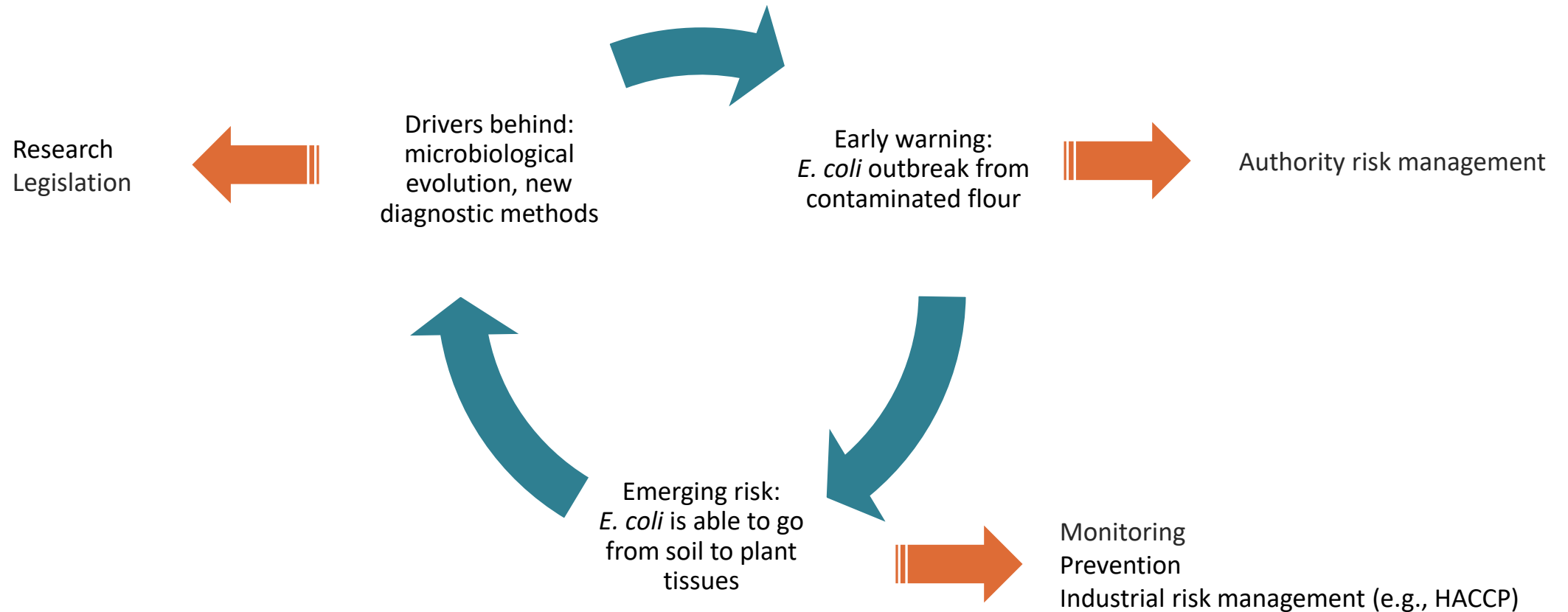


Foresight



- ▶ Long timescale studies → to induce thinking & actions on a strategic level
- ▶ Various techniques: driver analysis, scenario analysis, forecasting, back-casting, etc.
- ▶ **Foresight:** a systematic, participatory and multi-disciplinary approach to explore mid- to long-term futures and drivers of change
- ▶ **Drivers:** factors causing change, affecting or shaping the future.
- ▶ **Scenario analysis:** complex analysis of drivers, generating plausible future scenarios

EXAMPLE: transmission of *E. coli* from contaminated water and soil to plant tissues





Emerging risk identification

Emerging risk identification



COLLECTION AND COLLATION OF DATA & INFORMATION

- ▶ Screening various data & information sources
 - ▶ media and scientific literature
 - ▶ data from food safety authorities
 - ▶ patent databases
- ▶ Collecting expert knowledge



ANALYSIS AND FILTERING

- ▶ Characteristics assessed:
 - ▶ novelty, significance, susceptibility
- ▶ Prioritization: evaluation based on pre-defined criteria
 - ▶ soundness, imminence, scale, severity
 - ▶ risk management situation



SHARING

- ▶ Risk management/preparedness
 - ▶ data collection, monitoring, etc.
- ▶ Communication with stakeholders
- ▶ Research

Two approaches of data assisted Emerging risk identification

Imitating the human decision-making process

- ▶ Define what is 'emerging', what is 'risk', what is 'new technology', etc.
- ▶ Develop ontologies
- ▶ Develop search strings or automated data/information retrieval pipelines
- ▶ *Limitation/challenge: novelty is context-dependent*
- ▶ *LLMs might be useful here*

Pattern-based emergence definition

- ▶ Emergence is defined as a (significant) change in patterns/structures of a system
- ▶ Time-dynamics are important
- ▶ Systems approach
- ▶ *Many AI models, but still need human interpretation*
- ▶ *Needs a profound understanding of the algorithms behind*

Today

- ▶ Focus on 2 classes of methodologies of data-assisted emerging risk identification
- ▶ Both apply pattern-based, systemic approach
 - ▶ Prediction models
 - ▶ Text mining models



Funded by
the European Union

Topic modelling

Zsuzsa Farkas, Ákos Józwiak

University of Veterinary Medicine Budapest

HOLIFOOD Living Lab 1: Identification and monitoring of food safety risks

17 June 2024



Topic modelling

Overview

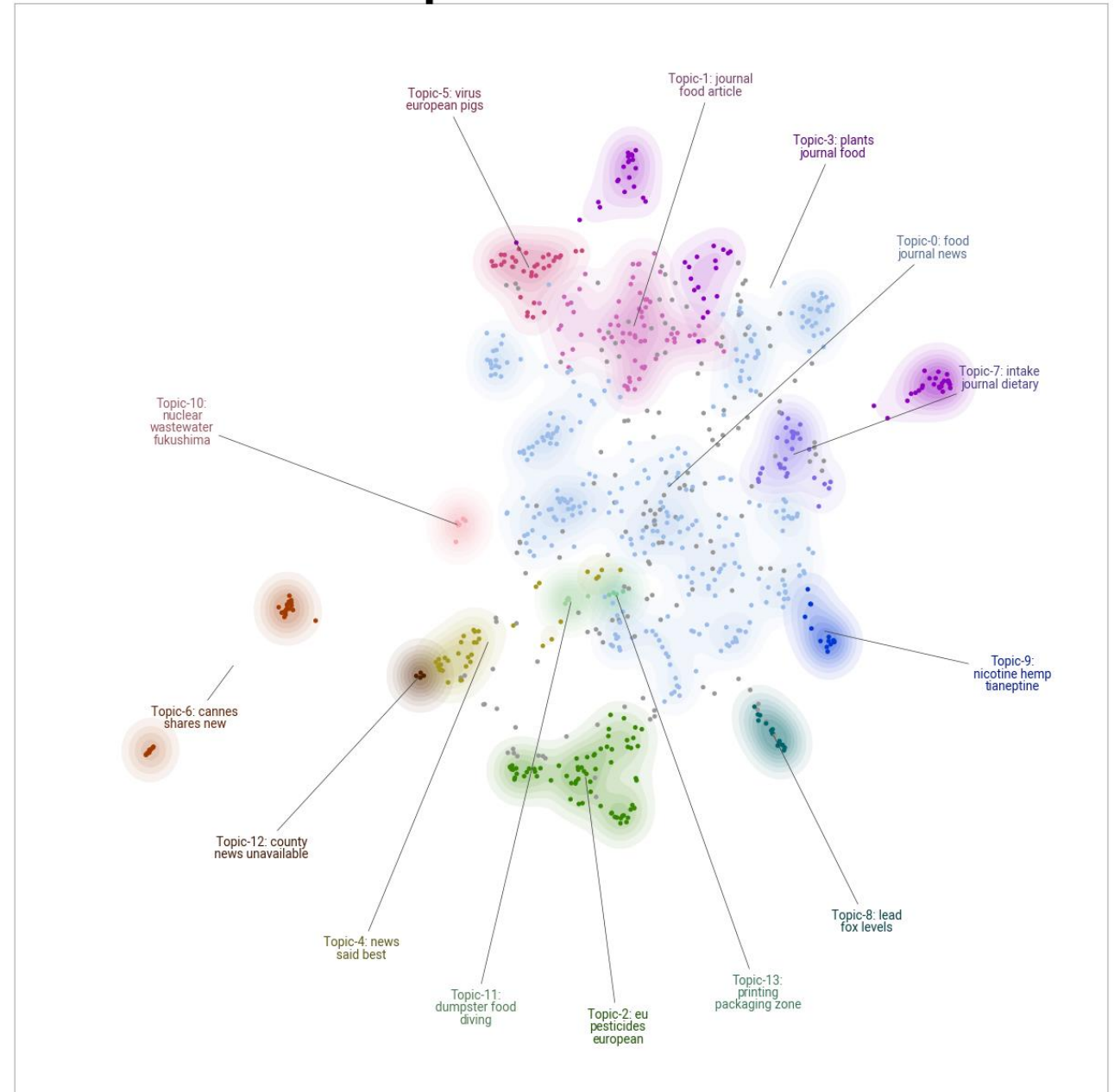
Identification of emerging risks from a corpus of text using AI

- ▶ Problem: information on emerging risks, or mechanisms leading to these already exist, but hard to find in the huge noise.
- ▶ Objective: to assist emerging risk identification experts with filtering and visualising various clusters of potential emerging issues
 - ▶ These methods will not tell what the emerging issues are, nor what the future will be
 - ▶ These methods are suitable for quick analysis of large text corpora, for decreasing the noise, and for separating 'important' things from 'less important' things
- ▶ Methods: raw signal miner, topic modelling

Topic modelling

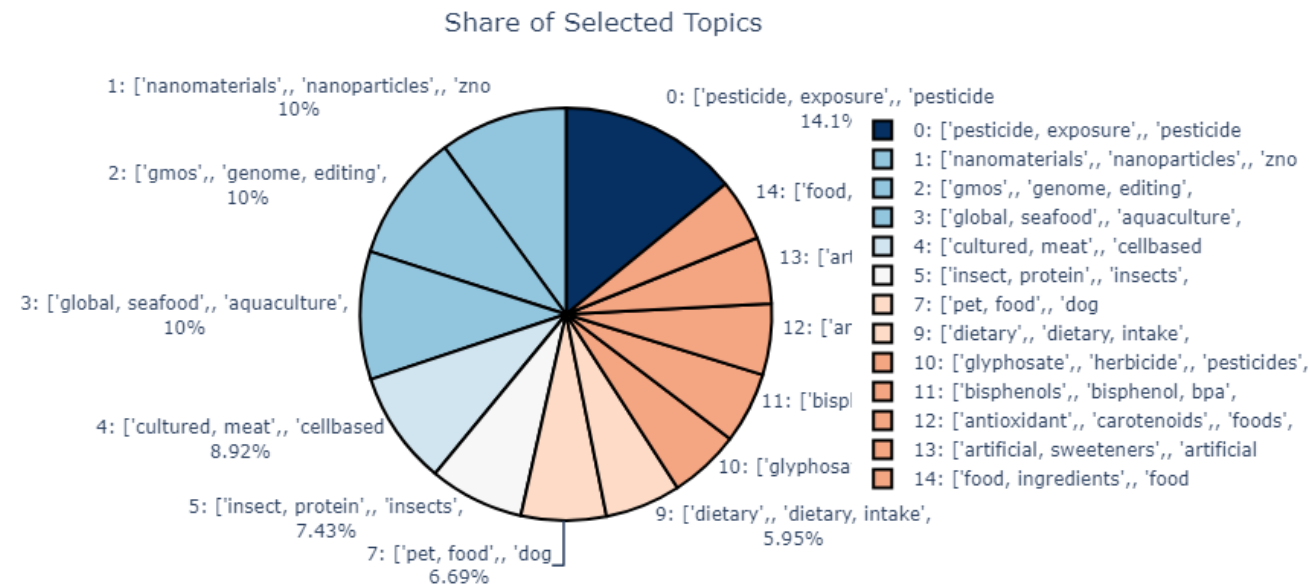
- What does the model do?
- How does the model work?
- Input (data sources)
- How is the model developed?
- Output (mock-up)

Documents and Topics



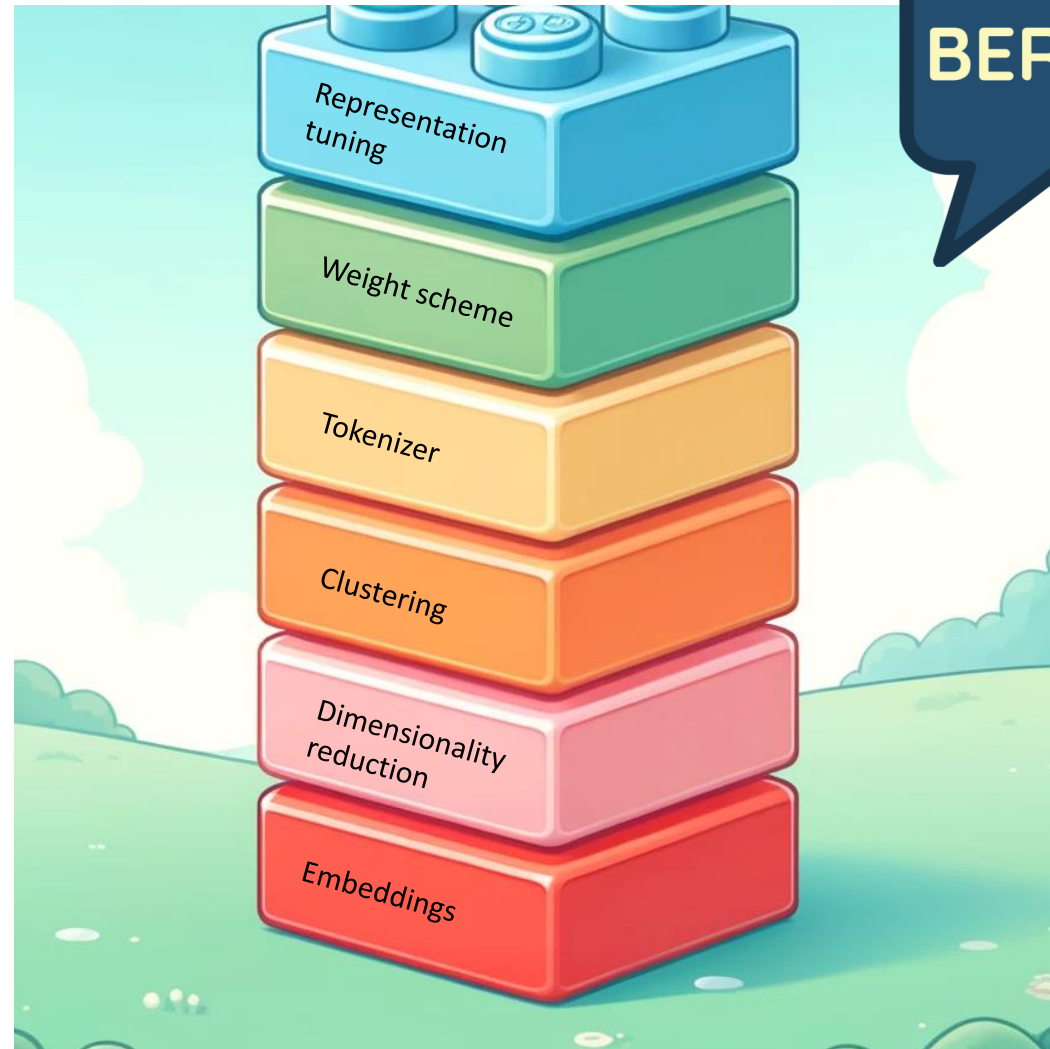
Topic modelling: what does the model do?

- ▶ Arranges textual data (e.g. documents, news articles) into topics with the help of text mining methodologies
- ▶ Unstructured textual data → Topics represented by keywords



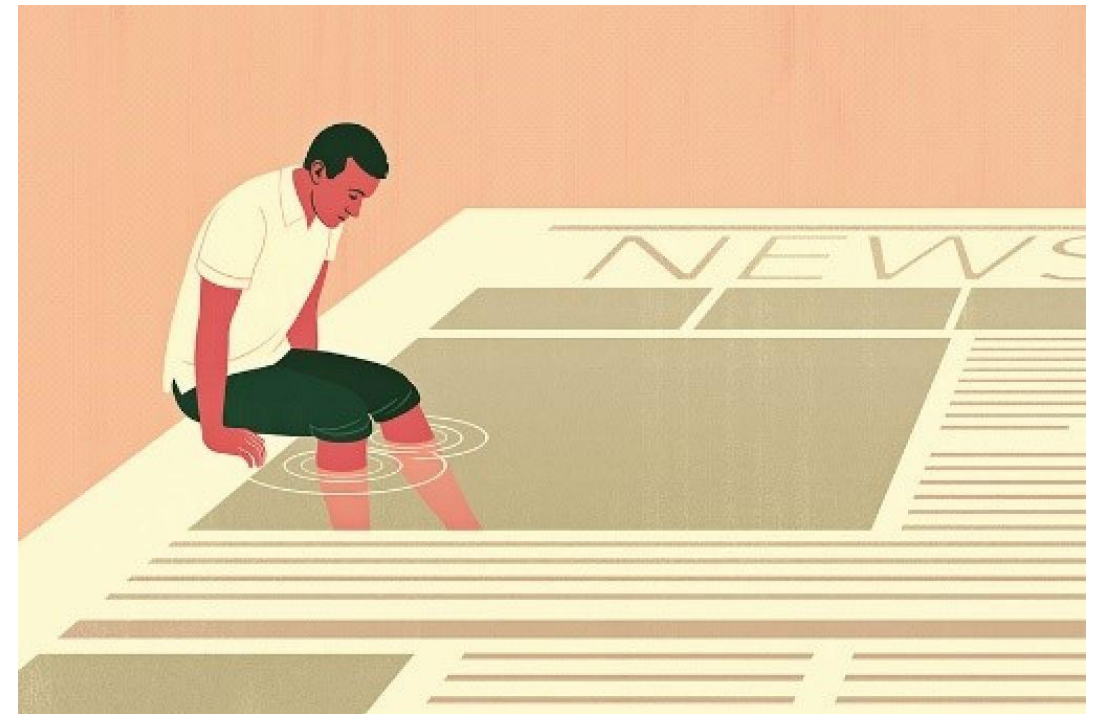
Topic modelling: how does the model work?

- ▶ BERTopic framework
 - ▶ Open source ML model
 - ▶ Python implementation
 - ▶ Focus on topic detection and representation
 - ▶ Modular design
 - ▶ Most advanced algorithms
 - ▶ c-TF-IDF for vectorization, BERT embedding
 - ▶ HDBSCAN – density based clustering
 - ▶ UMAP – dimensional reduction



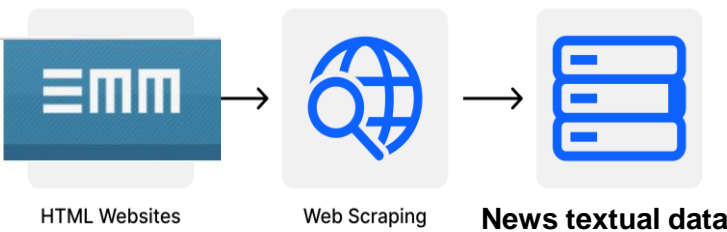
Topic modelling: input data

- ▶ Europe Media Monitor (EMM) – news that belong to food safety category
- ▶ Weekly retrieval of textual data by web scraping
- ▶ Translation from different languages

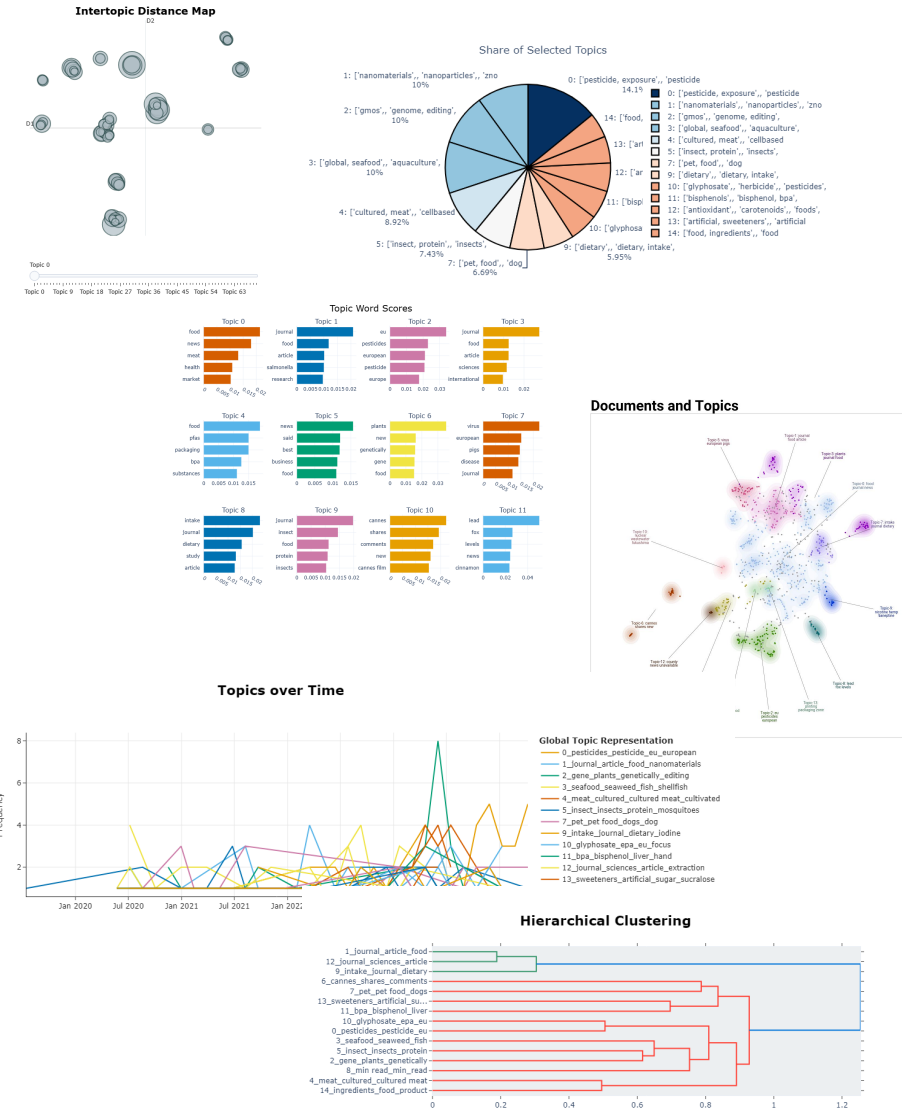


Topic modelling: how is the model developed?

Data retrieval

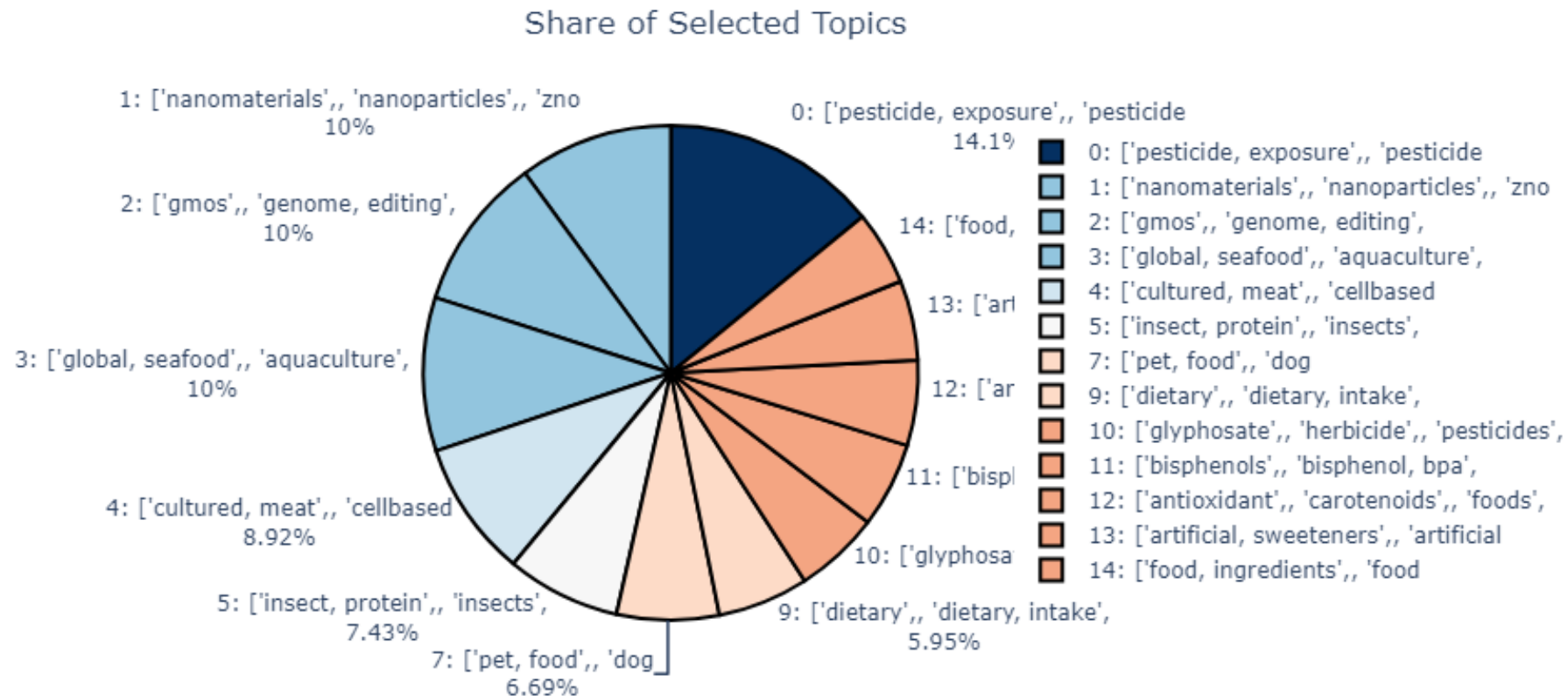


- ▶ Model training:
 - ▶ 3 years of EMM news
 - ▶ Manually selected (possible) emerging risks
 - ▶ Fine-tuning of parameters
- ▶ Weekly analysis of EMM news with the trained model



Topic modelling: outputs of the model

- Piechart of selected topics



Topic modelling: outputs of the model

- ▶ Barchart based on word scores

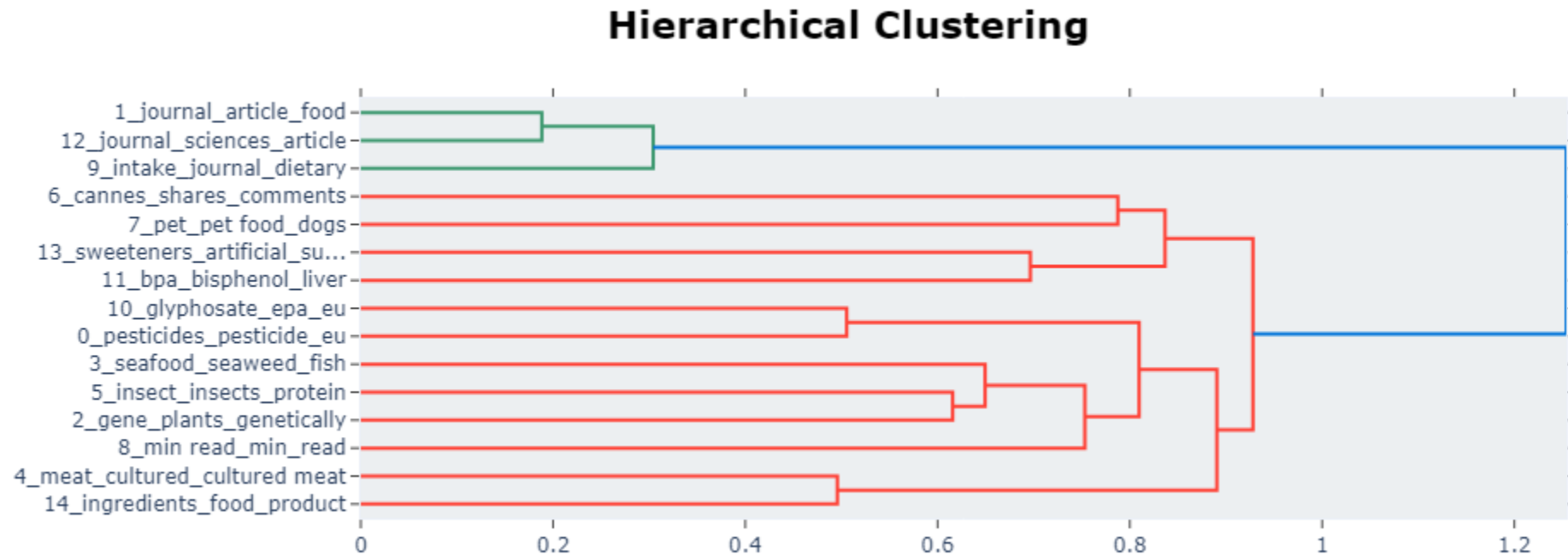


Topic modelling: outputs of the model

- Hierarchical clustering of topics

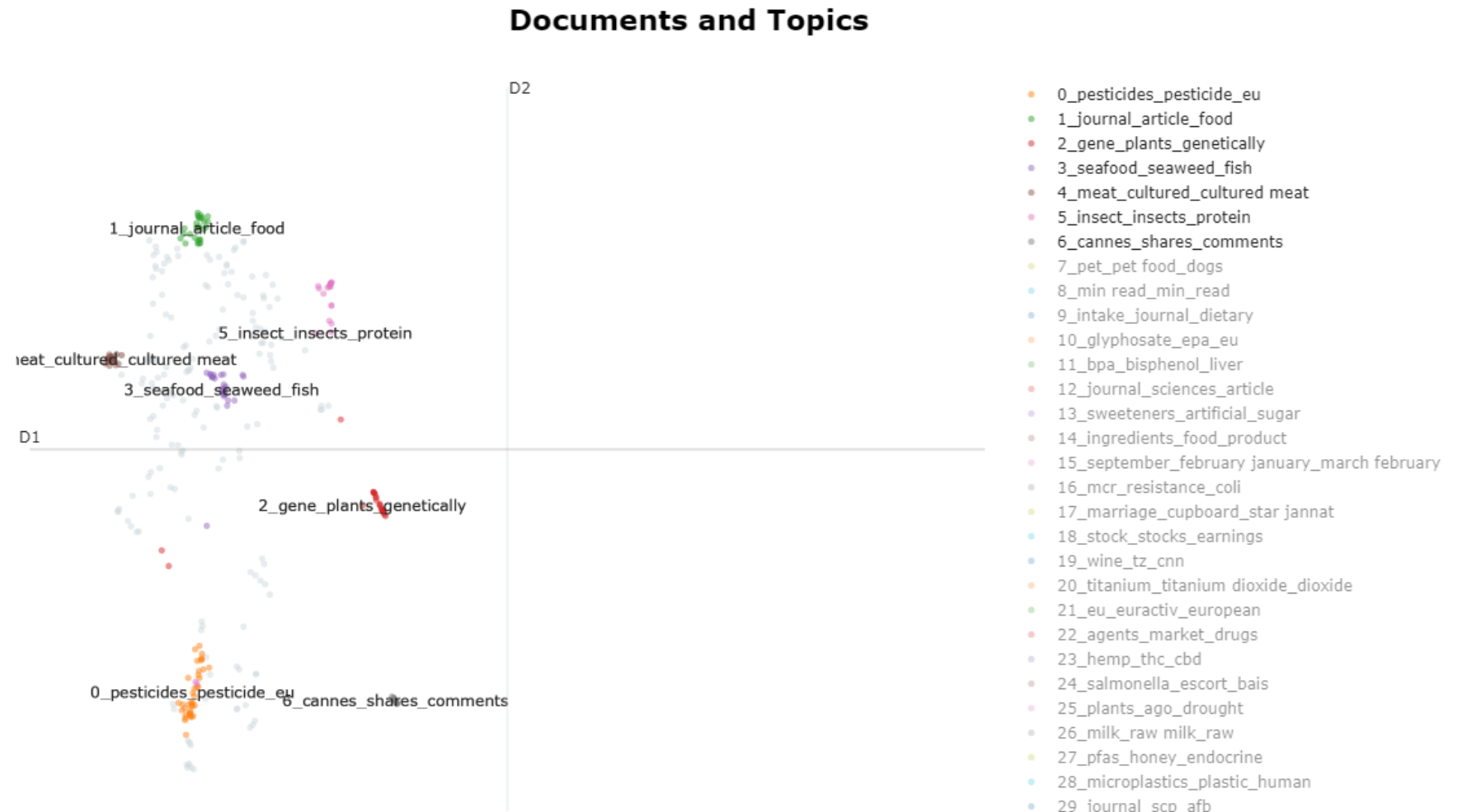


[hierarchy_visualization.html](#)



Topic modelling: outputs of the model

▶ 2D representation

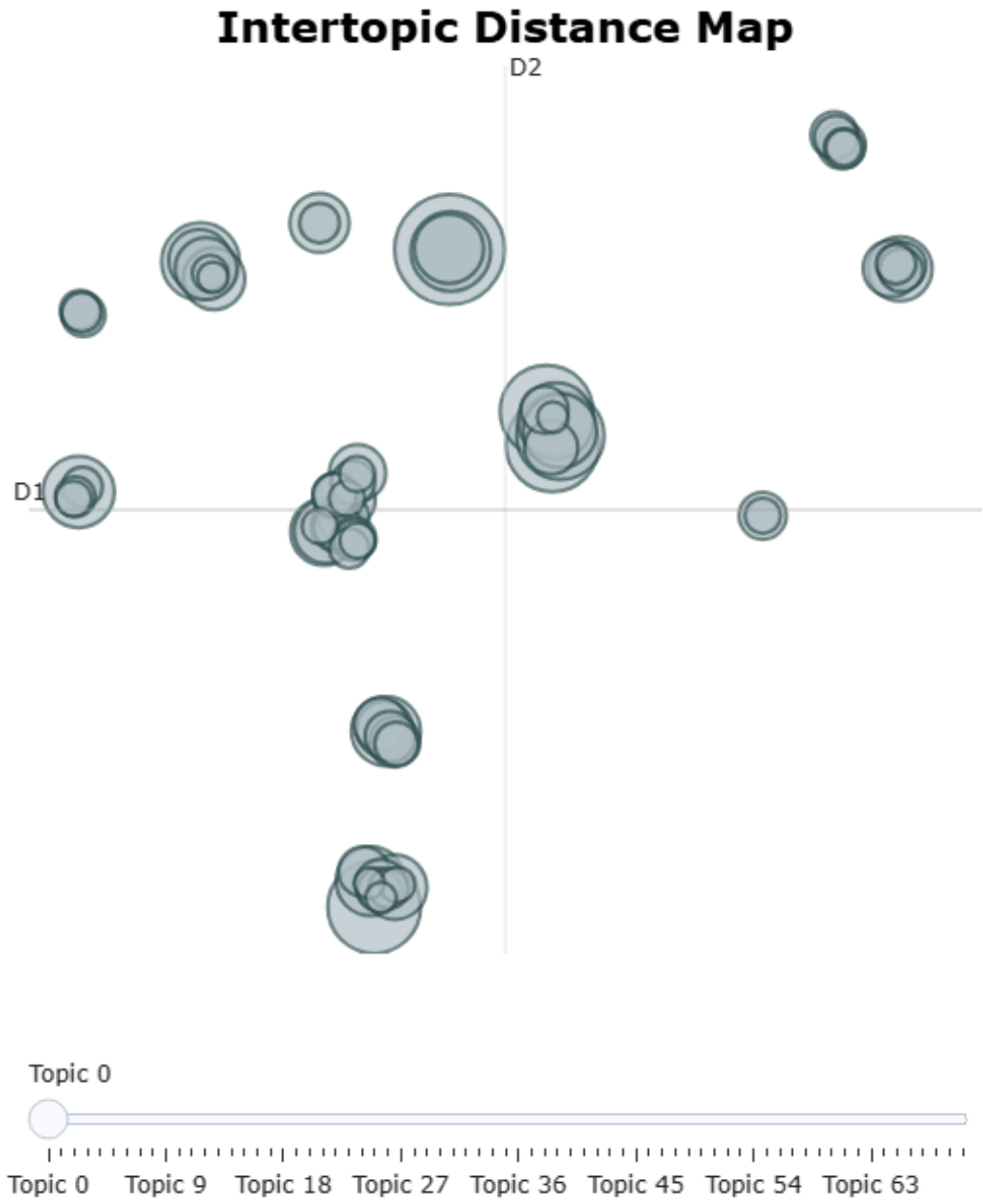


Topic modelling: outputs of the model

- ▶ Intertopic distance map



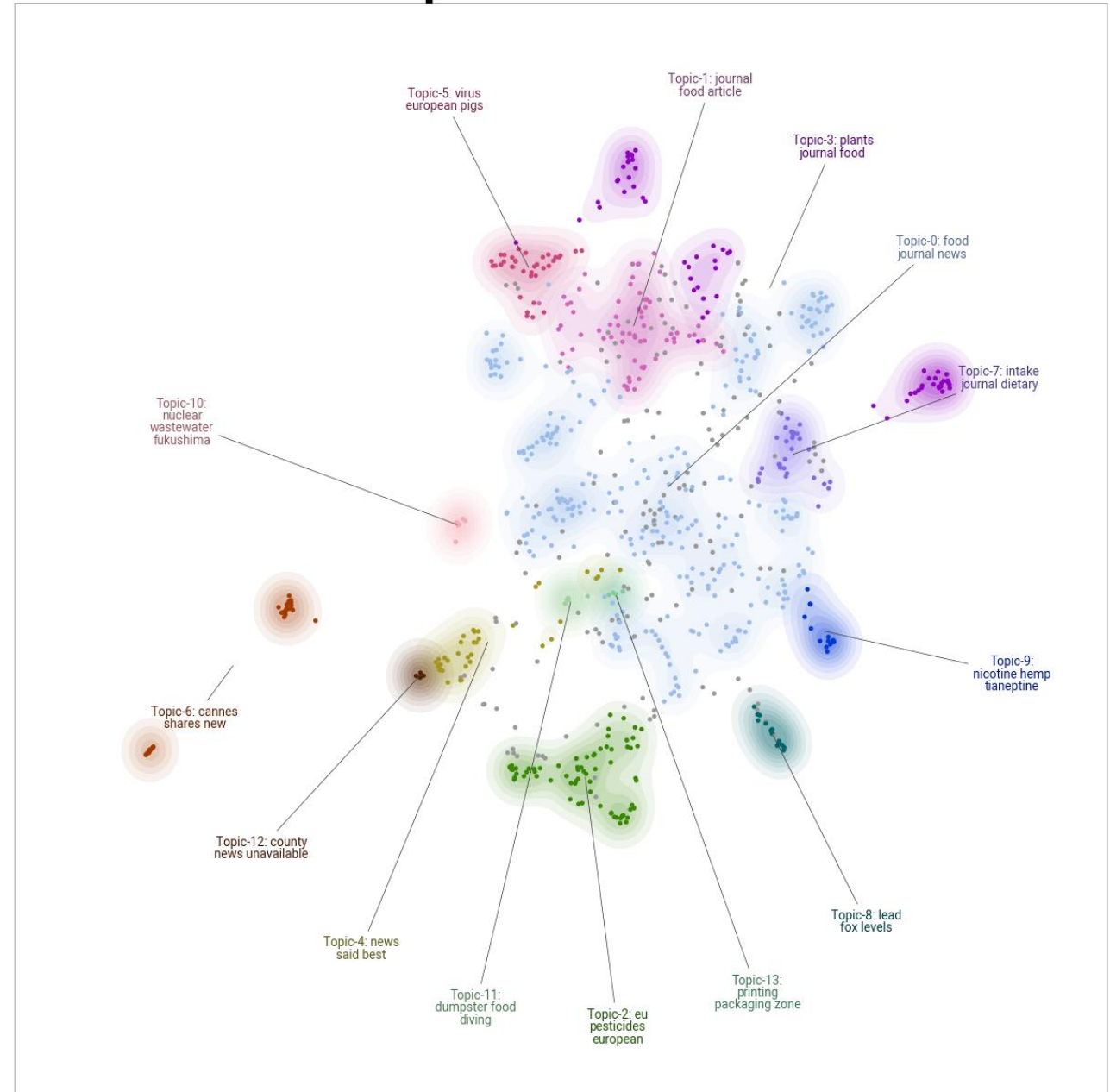
intertopics_visualization.html



Topic modelling: outputs of the model

- ▶ 3D representation
- ▶ QUESTION in the interactive session on !! RESOLUTION !!

Documents and Topics



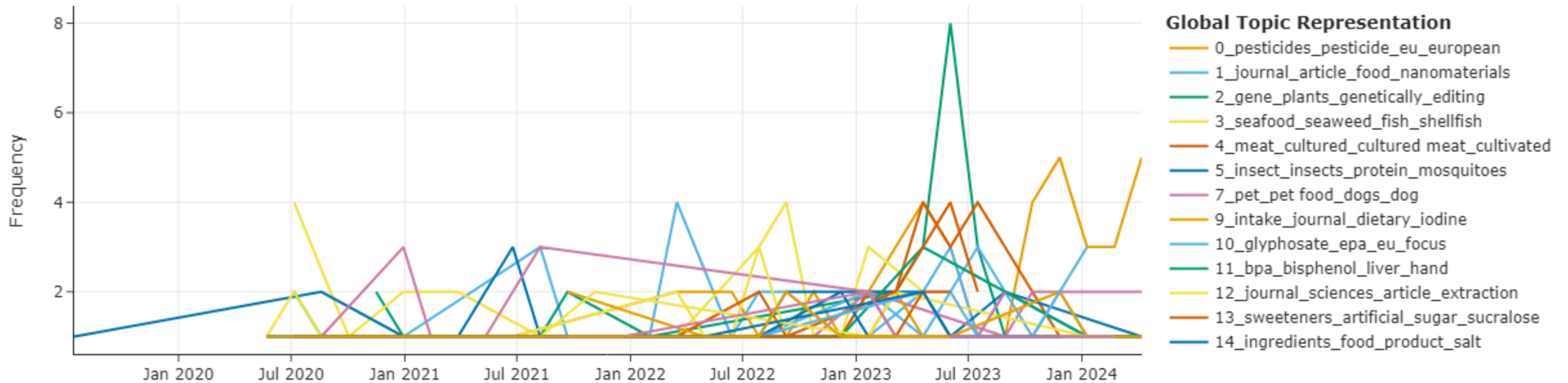
Topic modelling: outputs of the model

- ▶ Dynamic change of topics
- ▶ QUESTION in the interactive session on !! TIMEFRAME !!



dynamictopic_visualization.html

Topics over Time



Topic modelling: what is expected today?

- ▶ We need user input on the following questions:
 - ▶ Interpretability of outputs of the model: which of the visualizations do you find the most useful for emerging risk identification?
 - ▶ Timeframe of analysis: which timeframe is more relevant to you?
 - ▶ Resolution: how many topics should be presented at once?
- ▶ Method: Miro board
 - ▶ Showing pictures on the 3 main questions
 - ▶ Participants evaluate the visual outputs according to dimensions presented on Miro
 - ▶ Participants vote for the better visual for each of the 3 questions



Funded by
the European Union

HOLiFOOD Living Lab

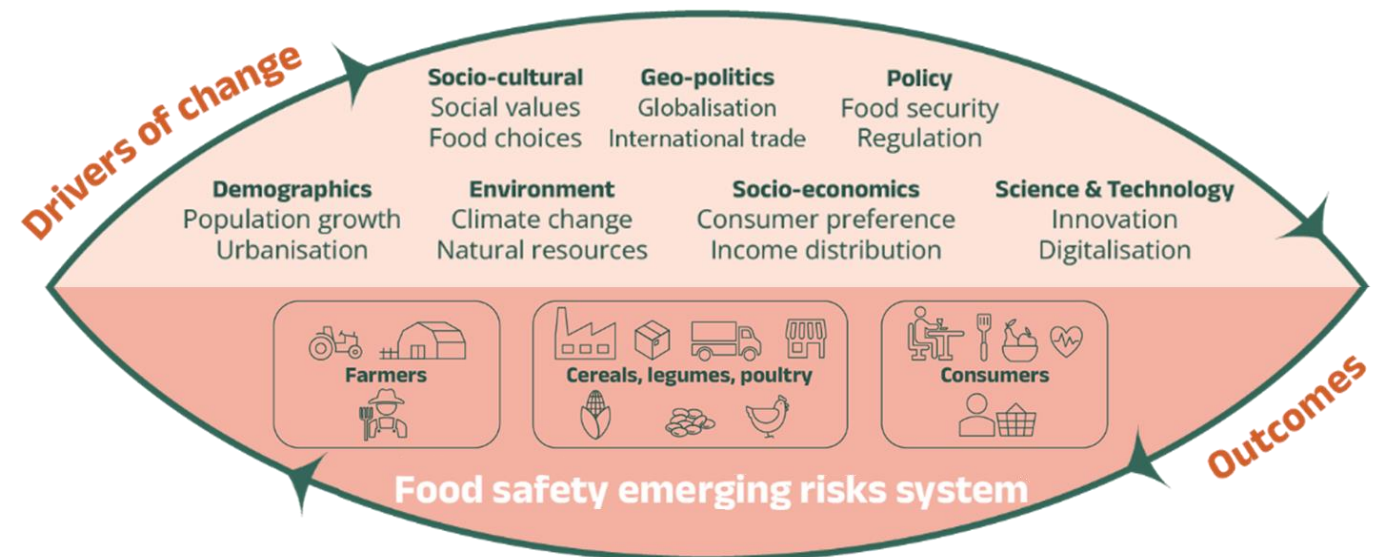
Prediction model breakout session

Leonieke van den Bulk (Wageningen Food Safety Research)

June 17th 2024

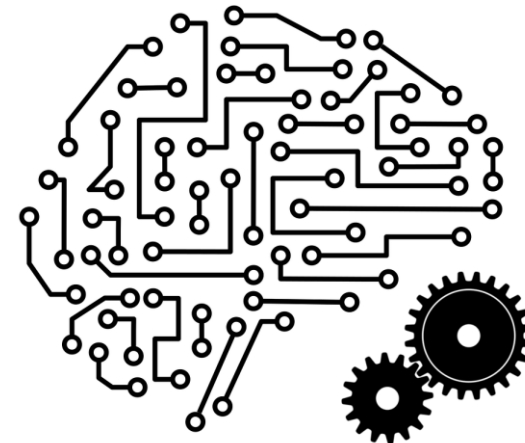
Prediction models – Emerging risks

- ▶ To identify **emerging risks** for food safety, we are developing prediction models to create a more proactive system
- ▶ Prediction of emerging risks is a **challenging problem** as food safety is driven by many factors, with direct and indirect influences
- ▶ We therefore need a **holistic approach**, which takes into account drivers of change across many domains

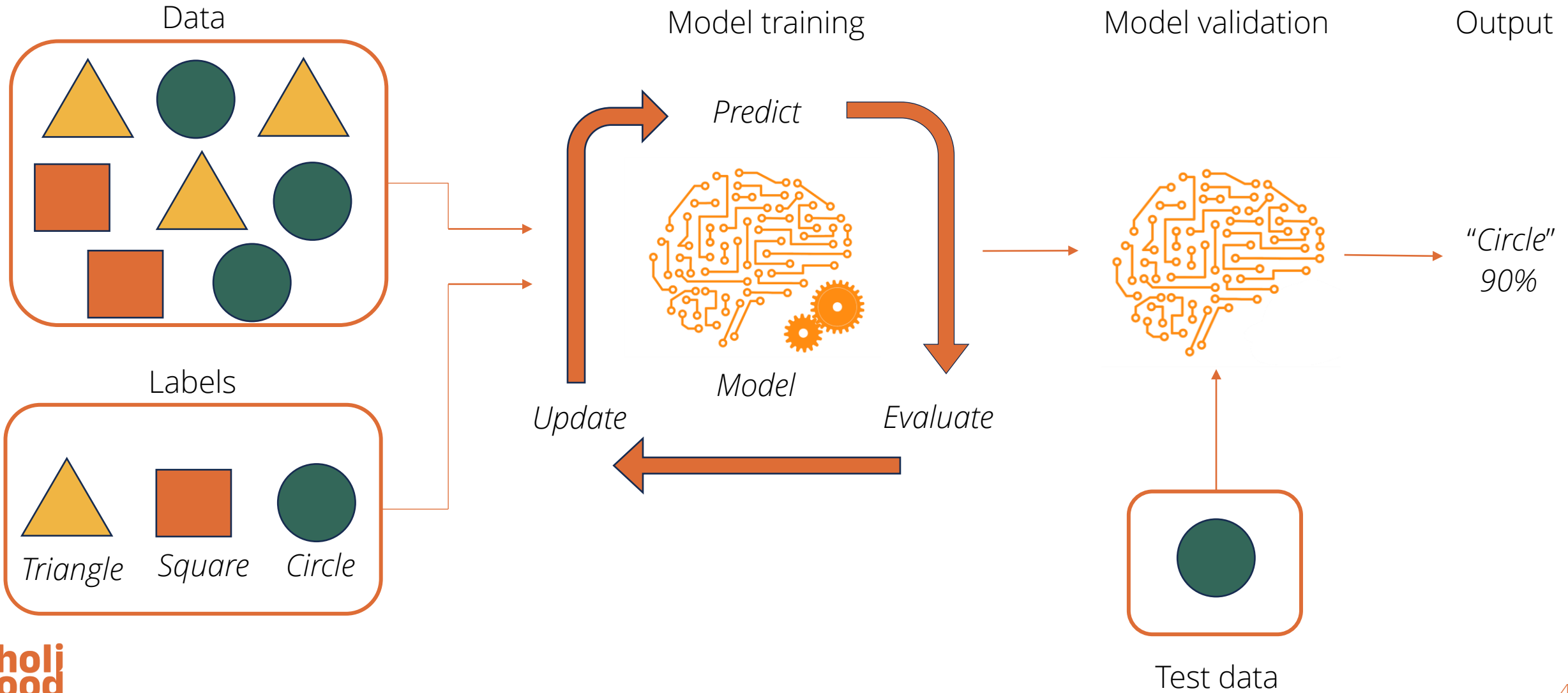


Prediction models – Machine learning

- ▶ We are building prediction models using the Artificial Intelligence technique: **machine learning**
- ▶ Machine learning encompasses algorithms/models that try to find **patterns** in data automatically
- ▶ This is done by trying to minimize the error between its **predictions** and the **true value** by updating its parameters



Prediction models - Flowchart



Prediction models - Data

- ▶ To make good models we need good data → EFSA monitoring data
- ▶ Contains monitoring data from **29** EU countries + the United Kingdom across many different products and contaminants with **millions** of measurements
- ▶ Information includes product, contaminant, concentration, origin country, sampling country, sampling data etc.



Prediction models – Use case



- ▶ Use case: **Pesticide** contamination of **lentils**
- ▶ In 2022, EFSA assessed that **3.7%** of food samples have pesticides residues above the legal limit ^[1]
- ▶ Pesticides are the contaminant most often measured > LOQ in lentils, with **1666 measurements > LOQ** out of a total of 942957 measurements

[1] European Food Safety Authority (EFSA), et al. "The 2022 European Union report on pesticide residues in food." EFSA Journal 22.4 (2024): e8753.

Prediction models - Drivers

- ▶ The selected **drivers of change** are important for model performance, since they should be able to add predictive power
- ▶ Drivers, selected by food safety experts, focus on:
 - **climate** (temperature, precipitation, humidity, disasters),
 - **economics** (prices, trade, GDP),
 - **social** (human development, press freedom),
 - **policy** (corruption, legal system, food security),
 - **technology** (innovation, digitalisation)
 - **production** (energy prices, pesticide usage)



Prediction models - Concept

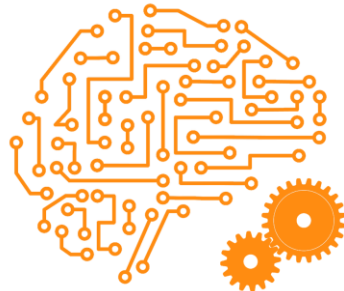
Monitoring data



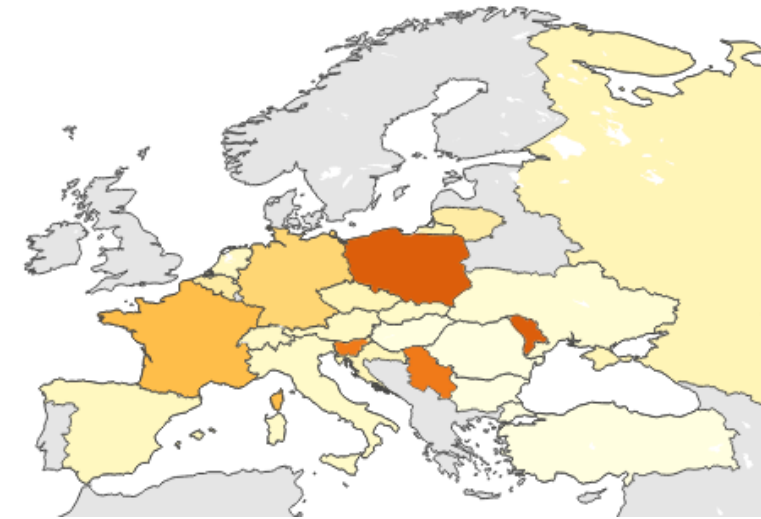
Drivers of change



Prediction model



Food safety contamination





Envisioned prediction tool

Predicted exceedance

Drivers of change

PRODUCT

lentils

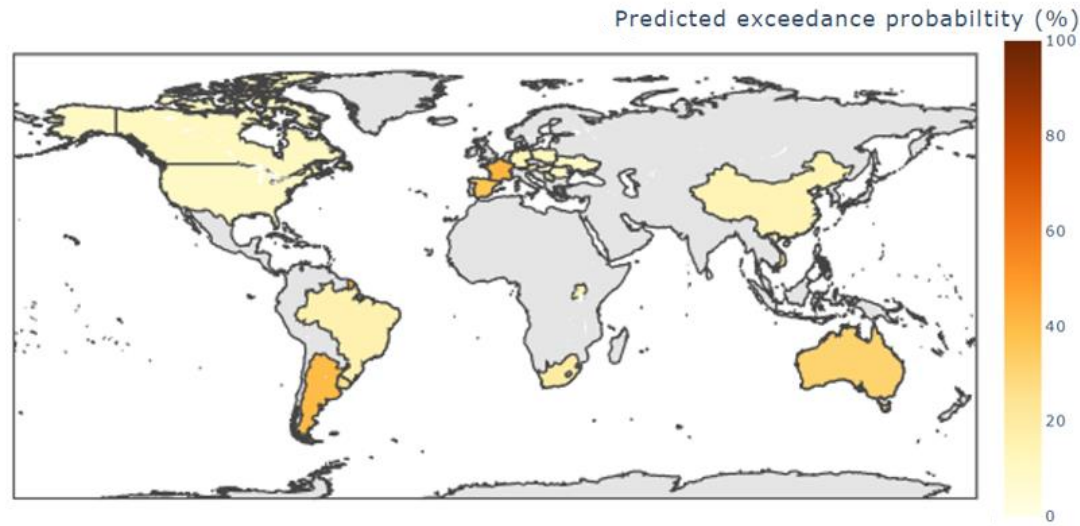
CONTAMINANT

pesticides

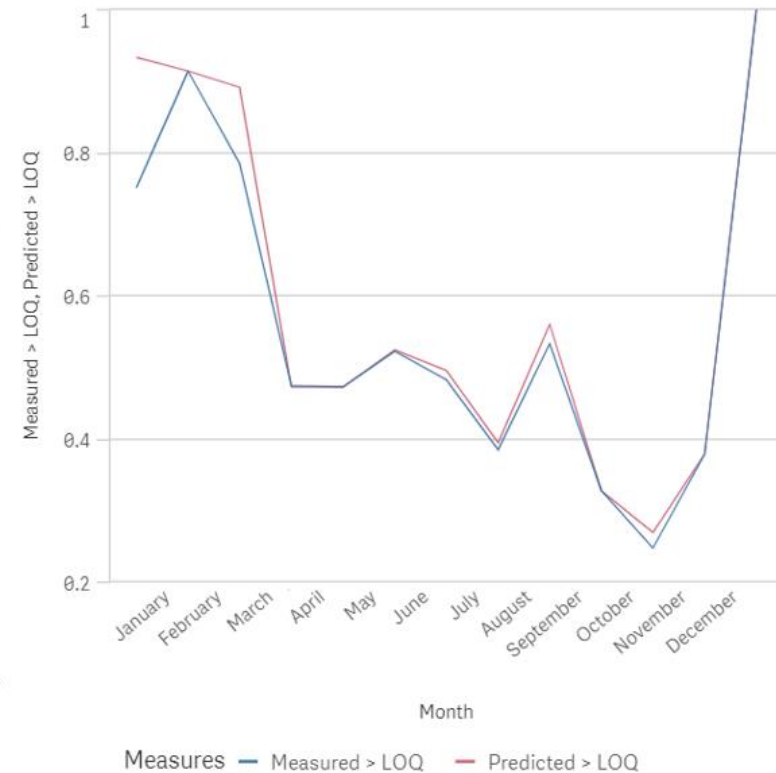
COMPUTE PREDICTION

PROBABILITY OF EXCEEDING THE LIMIT

COUNTRY OF ORIGIN	PREDICTED PROBABILITY (%)
France	41.5
Argentina	39.7
Spain	36.2
Australia	31.2
Uruguay	25.6
South Africa	19.3
Austria	17.9
Brazil	15.8
Uganda	15.1
Vietnam	14.6



VALIDATION OF EXCEEDANCE PROBABILITY



Predicted exceedance

Drivers of change

COUNTRY OF ORIGIN

France

PRODUCT

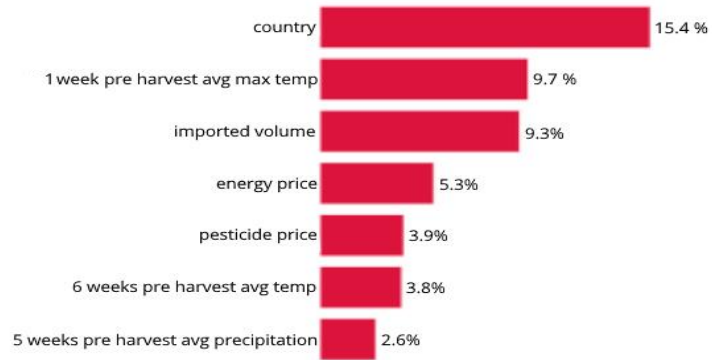
lentils

CONTAMINANT

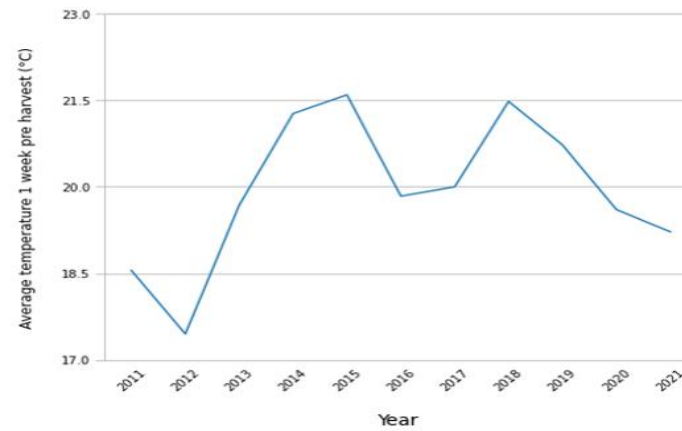
pesticides

COMPUTE DRIVERS

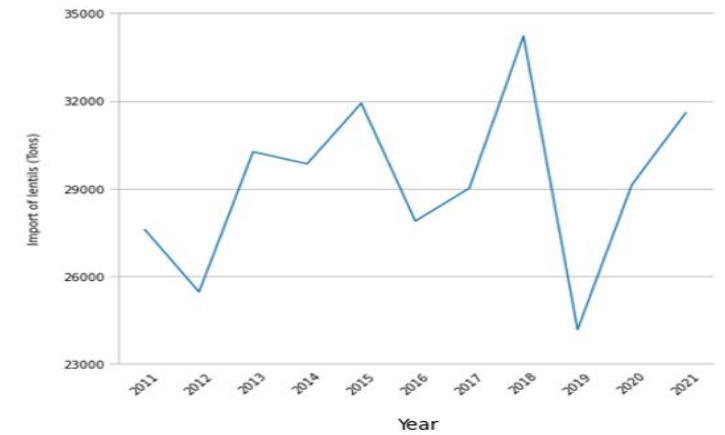
TOP PREDICTION INFLUENCING DRIVERS



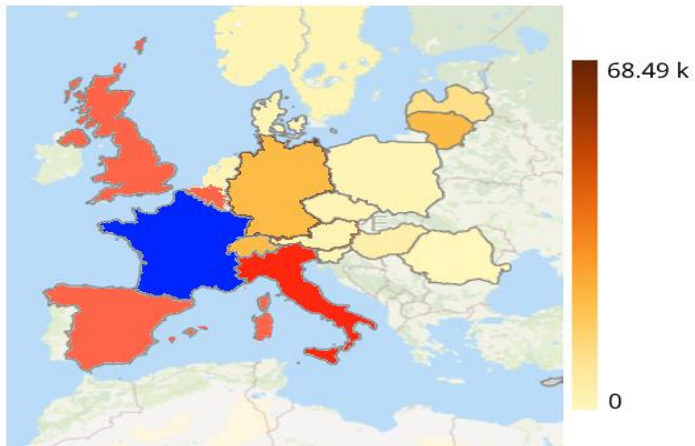
AVERAGE TEMPERATURE 1 WEEK PRE HARVEST



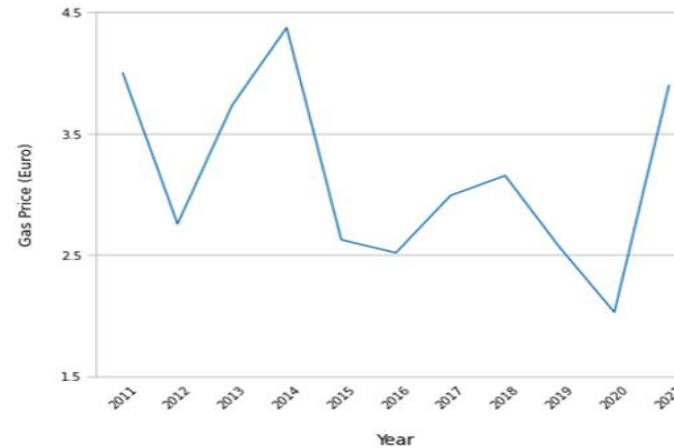
IMPORT OF LENTILS



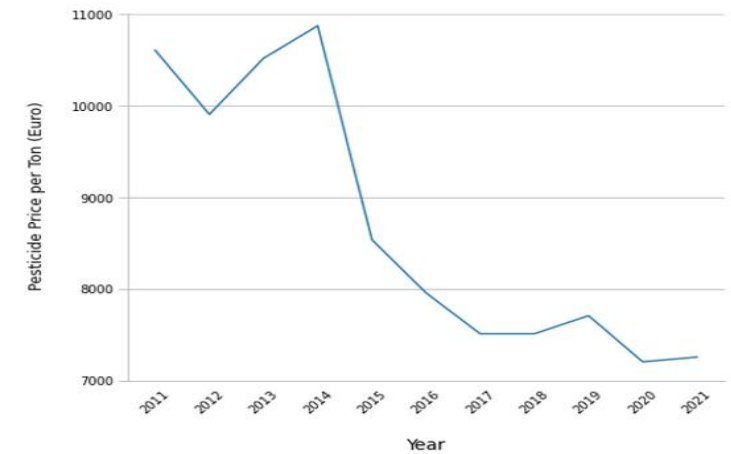
EXPORT OF LENTILS



GAS PRICE



PESTICIDE PRICE



Thank you!

Questions?



Funded by
the European Union