



# Extreme Food Risk Analytics

## D2.3.3: Report on Data Expansion and Population

Lead Authors: Juli Bakagianni (Agroknow) and  
Mihalis Papakonstantinou (Agroknow)



**Funded by  
the European Union**

Grant Agreement No.	101093026		
Project Acronym	EFRA		
Project Title	Extreme Food Risk Analytics		
Type of action	HORIZON Research and Innovation Actions		
Call Topic	HORIZON-CL4-2022-DATA-01-05		
Project Start Date	January 1 <sup>st</sup> , 2023	Project End Date	December 31 <sup>st</sup> , 2025
Project URL	efraproject.eu		
Work Package	WP2   Extreme Data Discovery & Mining		
WP Lead Beneficiary	Agroknow (AGK)		
Deliverable type <sup>1</sup>   Dissemination level <sup>2</sup>	Report   PU		
Contractual due date	March 31 <sup>st</sup> , 2024 (D2.3.1) December 31 <sup>st</sup> , 2024 (D2.3.2) December 31 <sup>st</sup> , 2025 (D2.3.3)	Actual submission date	March 29 <sup>th</sup> , 2024 (D2.3.1) January 31 <sup>st</sup> , 2025 (D2.3.2) January 16 <sup>th</sup> , 2026 (D2.3.3)
Lead Author (s)	Juli Bakagianni (Agroknow) and Mihalis Papakonstantinou (Agroknow)		
Contributors	Aron Henriksson (SU), Anna Romanova (SGS), Iraklis Posnakidis (Agroknow), Konstantinos Pechlivanis (Agroknow), Morgane Romeau (Agrivi), Ali Hurriyetoglu (WFSR), Alessio Bosca (Jakala), Franco Maria Nardini (CNR), Babis Thanopoulos (Agroknow)		
Internal reviewer(s)	Aron Henriksson (SU)		

### Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission-EU. Neither the European Union nor the granting authority can be held responsible for them.

### Copyright message

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

<sup>1</sup> R: Document, report; DEM: Demonstrator, pilot, prototype, plan designs; DEC: Websites, patents filing, press & media actions, videos, etc.; DATA: Data sets, microdata, etc; DMP: Data management plan; ETHICS: Deliverables related to ethics issues; SECURITY: Deliverables related to security issues; OTHER: Software, technical diagram, algorithms, models, etc.

<sup>2</sup> PU – Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page); SEN – Sensitive, limited under the conditions of the Grant Agreement; Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444; Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No2015/444; Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444



**Revision history (including peer reviewing & quality control)**

Version	Issue Date	% Complete	Changes	Contributor(s)
V2.1	30/06/2025	10%	Initial Deliverable Structure (ToC) – send for review	Juli Bakagianni (Agroknow) and Mihalis Papakonstantinou (Agroknow)
V2.2	11/07/2024	20%	Updated scope of the new version of the deliverable (D2.3.3) & methodology for data population monitoring	Juli Bakagianni (Agroknow)
V2.3	25/07/2024	30%	Allocation of chapter contribution to all involved partners	Juli Bakagianni (Agroknow)
V2.4	30/09/2024	75%	Content related to regulation datasets added by SGS	Anna Romanova (SGS), Franco Maria Nardini (CNR)
V2.5	31/10/2025	90%	Content related to EFRA Use cases added	Mihalis Papakonstantinou (Agroknow), Morgane Romeu (Agrivi), Anna Romanova (SGS)
V2.6	01/12/2025	100%	Content related to new datasets by SU, CNR & WFSR added	Aron Henriksson (SU), Ali Hurriyetoglu (WFSR), Franco Maria Nardini (CNR)
V2.7	18/10/2025	90%	Updates in all chapters in terms of responding to EFRA KPIs	Juli Bakagianni (Agroknow), Mihalis Papakonstantinou (Agroknow), Aron Henriksson (SU), Anna Romanova (SGS), Morgane Romeau (Agrivi), Ali Hurriyetoglu (WFSR), Alessio Bosca (Jakala), Franco Maria Nardini (CNR)
V2.8	22/10/2025	90%	Annex tables updated	Juli Bakagianni (Agroknow), Mihalis Papakonstantinou (Agroknow)
V2.9	09/01/2026	100%	Internal peer review	Aron Henriksson (SU)
<b>V3.0 (D2.3.3)</b>	<b>16/01/2026</b>	<b>100%</b>	<b>Addressed review comments and finalised deliverable</b>	<b>Juli Bakagianni (Agroknow), Babis Thanopoulos (Agroknow)</b>

EFRA Consortium			
#	Participant Organisation Number	Short name	Country
1	AGROKNOW IKE	AGROKNOW	EL
2	CONSIGLIO NAZIONALE DELLE RICERCHE	CSR	IT
3	STOCKHOLMS UNIVERSITET	SU	SE
4	STICHTING WAGENINGEN RESEARCH	WR	NL
5	JAKALA S.p.A. S.B.	JAKALA	IT
6	AGRIVI DOO ZA PROIZVODNJU TRGOVINUI USLUGE	AGRIVI DOO	HR
7	RAINNO IDIOTIKI KEFALAIOUCHIKI ETAIREIA	RAINNO	EL
8	SGS ROMANIA SA	SGS ROMANIA	RO
9	MOY PARK LTD	MOY PARK	UK

## Table of Contents

<b>1.</b>	<b><i>Executive Summary</i></b>	<b>9</b>
<b>2.</b>	<b><i>Introduction</i></b>	<b>10</b>
2.1.	Mapping EFRA Outputs	10
2.2.	Deliverable Overview and Structure	10
<b>3.</b>	<b><i>Data Population Methodology</i></b>	<b>12</b>
3.1.	Overview and Scope	12
3.2.	Data Registry Tool for Data Population and Monitoring	12
3.3.	Monitoring, Coordination, and Progress Tracking During RP2	12
3.4.	KPI Monitoring Framework: Definitions and Scope	12
3.5.	Link to Per - Data Source Reporting	13
<b>4.</b>	<b><i>Composite Datasets</i></b>	<b>14</b>
4.1.	SGS Digicomply Dataset	14
4.2.	Chunk Relevance Validation Dataset	15
4.3.	The FoodSafeSum Dataset	16
<b>5.</b>	<b><i>Project-Generated Data Sources</i></b>	<b>19</b>
5.1.	Food Recall Incidents	19
5.2.	Scientific Publications	20
5.2.1.	Scientific Publications from FoodSafeSum Dataset	20
5.2.2.	Abstracts of a Systematic Review on Food and Federated Learning	21
5.2.3.	Abstracts of a Systematic Review on Food and NLP	22
5.3.	News Articles	23
5.3.1.	News Articles from FoodSafeSum Dataset	23
5.3.2.	News Articles by WFSR	23
5.4.	Governmental Guidance from FoodSafeSum Dataset	24
5.5.	Regulations from FoodSafeSum Dataset	24
<b>6.</b>	<b><i>Decision-Support / Pilot-Contributing Data Sources</i></b>	<b>25</b>
6.1.	Use Case 1 - Data Sources for the EFRA Decision Support Tool	25
6.1.1.	Food Product Recalls	25
6.1.2.	Global Food Product Border Rejections	26
6.1.3.	Food Safety News	26
6.1.4.	Foodborne Disease Outbreaks	27
6.1.5.	Inspections	28
6.1.6.	Lab Tests	29
6.1.7.	Weather Data	29
6.1.8.	Prediction Trends	30
6.1.9.	Prediction Risks	31
6.1.10.	Production Data	32
6.1.11.	Trade Data	33
6.1.12.	Price Data	33
6.1.13.	Poultry Production Data	34
6.1.14.	Food Fortress Dataset	35
6.1.15.	Global Fraud Reports	36
6.2.	Use Case 2 - Data Sources for the EFRA Decision Support Tool	37

6.2.1.	Weather Data	37
6.2.2.	Pest Observations	38
<b>6.3.</b>	<b>Use Case 3 - Data Sources for the EFRA Decision Support Tool</b>	<b>41</b>
6.3.1.	Scientific Publications	41
6.3.2.	News Articles	42
6.3.3.	Governmental Guidance	42
6.3.4.	Regulations	43
<b>7.</b>	<b>Collaborative / External Data Sources</b>	<b>45</b>
7.1.	Contextualized Social Recommender System Data Source (CoSRec)	45
7.2.	The CompreHensive European Food Safety (CHEFS) Database	47
<b>8.</b>	<b>Data Assets</b>	<b>50</b>
<b>9.</b>	<b>Conclusions</b>	<b>54</b>
<b>Annex I</b>		<b>55</b>
<b>Annex II</b>		<b>85</b>

## List of Tables

<b>Table 1:</b>	<b>Adherence to EFRA GA Deliverable &amp; Tasks Descriptions</b>	<b>10</b>
<b>Table 2:</b>	<b>Contribution of the SGS Digicomply dataset to the data source related KPIs</b>	<b>15</b>
<b>Table 3:</b>	<b>Contribution of the Chunk Relevance Validation dataset to the data source-related KPIs</b>	<b>16</b>
<b>Table 4:</b>	<b>Contribution of the FoodSafeSum dataset to the data source-related KPIs</b>	<b>18</b>
<b>Table 5:</b>	<b>Contribution of the Food Recall Incidents dataset to the data source related KPIs</b>	<b>20</b>
<b>Table 6:</b>	<b>Contribution of the scientific publications from FoodSafeSum dataset to the data source related KPIs</b>	<b>20</b>
<b>Table 7:</b>	<b>Contribution of the abstracts of a systematic review on food and federated learning to the data source-related KPIs</b>	<b>22</b>
<b>Table 8:</b>	<b>Contribution of the abstracts of a systematic review on food and NLP to the data source related KPIs</b>	<b>22</b>
<b>Table 9:</b>	<b>Contribution of the news articles from FoodSafeSum dataset to the data source related KPIs</b>	<b>23</b>
<b>Table 10:</b>	<b>Contribution of the news articles by WFSR to the data source related KPIs</b>	<b>24</b>
<b>Table 11:</b>	<b>Contribution of the governmental guidance from FoodSafeSum dataset to the data source related KPIs</b>	<b>24</b>
<b>Table 12:</b>	<b>Contribution of the regulations from FoodSafeSum dataset to the data source related KPIs</b>	<b>24</b>
<b>Table 13:</b>	<b>Contribution of the food product recalls to the data source related KPIs</b>	<b>25</b>
<b>Table 14:</b>	<b>Contribution of the global food product border rejections to the data source related KPIs</b>	<b>26</b>
<b>Table 15:</b>	<b>Contribution of the food safety news to the data source related KPIs</b>	<b>27</b>
<b>Table 16:</b>	<b>Contribution of the foodborne disease outbreaks to the data source related KPIs</b>	<b>28</b>
<b>Table 17:</b>	<b>Contribution of the inspections to the data source related KPIs</b>	<b>28</b>
<b>Table 18:</b>	<b>Contribution of the lab tests to the data source related KPIs</b>	<b>29</b>
<b>Table 19:</b>	<b>Contribution of the weather data to the data source related KPIs</b>	<b>30</b>
<b>Table 20:</b>	<b>Contribution of the prediction trends to the data source related KPIs</b>	<b>31</b>
<b>Table 21:</b>	<b>Contribution of the prediction risks to the data source related KPIs</b>	<b>32</b>
<b>Table 22:</b>	<b>Contribution of the production data to the data source related KPIs</b>	<b>32</b>
<b>Table 23:</b>	<b>Contribution of the trade data to the data source related KPIs</b>	<b>33</b>
<b>Table 24:</b>	<b>Contribution of the price data to the data source related KPIs</b>	<b>34</b>
<b>Table 25:</b>	<b>Contribution of the Moy Park poultry production data to the data source related KPIs</b>	<b>35</b>
<b>Table 26:</b>	<b>Contribution of the food fortress dataset to the data source related KPIs</b>	<b>36</b>
<b>Table 27:</b>	<b>Contribution of the fraud reports to the data source related KPIs</b>	<b>37</b>
<b>Table 28:</b>	<b>Contribution of the weather data by Agrivi to the data source related KPIs</b>	<b>38</b>
<b>Table 29:</b>	<b>Contribution of the pest observations to the data source related KPIs</b>	<b>41</b>
<b>Table 30:</b>	<b>Contribution of the scientific publications from SGS Digicomply Dataset to the data source related KPIs</b>	<b>41</b>
<b>Table 31:</b>	<b>Contribution of the Chunk Relevance Validation dataset to the data source related KPIs</b>	<b>41</b>
<b>Table 32:</b>	<b>Contribution of the news articles from SGS Digicomply dataset to the data source related KPIs</b>	<b>42</b>
<b>Table 33:</b>	<b>Contribution of the news articles of the Chunk Relevance Validation dataset to the data source related KPIs</b>	<b>42</b>
<b>Table 34:</b>	<b>Contribution of the governmental guidance from SGS Digicomply dataset to the data source related KPIs</b>	<b>43</b>

<b>Table 35:</b> Contribution of the governmental guidance from Chunk Relevance Validation dataset to the data source related KPIs.....	43
<b>Table 36:</b> Contribution of the regulations from SGS Digicomply dataset to the data source related KPIs.....	44
<b>Table 37:</b> Contribution of the regulations from Chunk Relevance Validation dataset to the data source related KPIs ...	44
<b>Table 38:</b> Contribution of the CosRec dataset to the data source related KPIs .....	47
<b>Table 39:</b> Contribution of the CHEFS database to the data source related KPIs.....	49
<b>Table 40:</b> EFRA Data Assets (name, data owner, number of records).....	50
<b>Table 41:</b> Aggregated KPI values per data source, including totals, targets, and differences.....	51
<b>Table 42:</b> The data types of the data sources provided by the owners that support KPI 1.1. ....	55
<b>Table 43:</b> The sources of the data sources provided by the owners that support KPI 1.2 .....	56
<b>Table 44:</b> The languages supported by the datasets provided by the owners to support KPI 1.11.....	82

## List of Figures

<b>Figure 1:</b> Screenshots of the EFRA Data Registry used for monitoring data population .....	85
--	----

## Glossary of terms and abbreviations used

Abbreviations / Acronyms	Meaning
KPI	Key Performance Indicator
RP	Reporting Period
Aggr. Data Vol.	Aggregated Data Volume
Soc. & News Data	Social & News Data
Sci. & Expert Knowl.	Scientific & Expert Knowledge
Env. & Clim. Data	Environmental & Climate Data
Econ. & Trade Data	Economic & Trade Data
Reg. & Legis. Data	Regulatory & Legislative Data
Sust. & Animal Health Data	Sustainability & Animal Health Data
Soc. & News Data	Social & News Data
Sci. & Expert Knowl.	Scientific & Expert Knowledge
Langs Supp.	Languages Supported
NLP	Natural Language Processing
LLM	Large Language Model
LOD	Limit of Detection Value
LOQ	Limit of Qualification Value



## 1. Executive Summary

This deliverable documents the data sources integrated, curated, and assessed within the EFRA project to support the development of AI-based food safety and decision-support services. It focuses on identifying, documenting, and integrating heterogeneous data sources relevant to food safety, and on demonstrating their analytical readiness across the EFRA ecosystem.

The work carried out includes the systematic cataloguing of project-related, decision-support, and collaborative or external data sources, covering a wide range of document types and data modalities such as scientific publications, regulations, governmental guidance, news articles, environmental records, economic indicators, and sustainability-related datasets. Each data source is documented in a structured manner, detailing its scope, content, provenance, data volume, accessibility, and contribution to the project's Key Performance Indicators (KPIs).

The deliverable also documents composite datasets, such as FoodSafeSum, which are curated resources that aggregate multiple individual data sources into a unified dataset for specific analytical purposes. These are presented alongside their constituent data sources, ensuring transparency and traceability while avoiding redundancy.

In this context, deliverable **D2.3.3** (3rd iteration) on **EFRA Data Population** provides an updated and consolidated overview of the EFRA data landscape. During this period, millions of records from public, industrial, collaborative, and project-generated sources have been examined, reflecting the significant expansion of EFRA's data ecosystem.

A substantial part of the work focuses on project-generated data sources, created directly by EFRA partners through internal research, systematic reviews, annotation activities, and targeted data collection. These include food recall incidents, scientific publications and systematic review abstracts, news articles, governmental guidance, and regulatory documents. They provide high-quality, domain-specific content that supports model development, relevance validation, and thematic analyses.

Complementing these are the decision-support and pilot-contributing data sources, which directly support EFRA's demonstrators, predictive models, and use cases. These datasets provide the contextual, environmental, scientific, and regulatory signals required for risk prediction, trend analysis, and decision-support outputs. They include, among others, lab tests, food incident recalls, weather data, scientific publications, news articles, governmental guidance, regulations, pest observations, inspections, prediction risks and trends, production and trade data, price data, and poultry production data. These resources are essential for the operation of EFRA's analytical pipelines and the development of the project's decision-support tools.

Across all categories, the deliverable highlights the heterogeneity of formats, linguistic diversity, and structural inconsistencies inherent in global food safety data. It outlines the validation and transformation workflows applied to convert these disparate sources into coherent, interoperable datasets suitable for EFRA's analytical and modelling needs. Overall, D2.3.3 demonstrates the breadth, depth, and maturity of EFRA's data landscape, showcasing its capacity to support advanced analytics, machine learning models, and decision-support tools for stakeholders across the food safety ecosystem.

## 2. Introduction

The purpose of this deliverable is to document the data assets considered within EFRA, outline their characteristics, and describe how they contribute to the project's data-related KPIs. In accordance with the GA, the deliverable provides a structured overview of all composite datasets and individual data sources examined during this reporting period, together with the processing workflows, metadata structures, and curation practices applied to them.

### 2.1. Mapping EFRA Outputs

**Table 1** summarises the alignment of this deliverable with the relevant EFRA tasks and deliverable requirements. It clarifies which tasks contributed to the dataset registry, processing pipelines, and KPI reporting.

**Table 1:** Adherence to EFRA GA Deliverable & Tasks Descriptions

EFRA GA Component Title	EFRA GA Component Outline	Respective Document Chapter(s)	Justification
Task 2.1 EFRA Data Hub population & heterogeneous data source registry	<p>T2.1 Description: This task will leverage the output of T1.2 to create an extensible registry of data sources &amp; types and relevant rules/specs/ algorithms to crawl, mine and aggregate them. The novel approach for automatically identifying promising and/or deep/ hidden food safety data sources will also be deployed and tested. By deploying the relevant technological stack over the WP4 Cloud HPC infrastructure, the EFRA Data Hub will be established and continuously enhanced and populated, providing the data to support all project activities.</p> <p>D2.3 Description: Report on the deployment of T2.1 over the identified data sources and the data types and data populations reached.</p>	Chapters 3, 4, 5, 6, 7, 8, Annex I, II	Chapters 3–4 cover methodology and composite datasets; Chapters 5–7 detail individual data sources; Chapter 8 summarizes data assets and KPI contributions. Annex I presents KPI calculations, and Annex II provides the central data registry snapshot. This coverage fulfils task objectives, including processing, metadata, and KPI documentation.

### 2.2. Deliverable Overview and Structure

The rest of this document is structured as follows:

- **Chapter 3, *Data Population Methodology*:** Describes the methodology used during RP2 for systematic population, monitoring, and documentation of data sources, ensuring coordinated acquisition and transparent reporting of KPI contributions.
- **Chapter 4, *Composite Datasets*:** Presents datasets that aggregate multiple data source types into unified resources, designed to support EFRA's analytical and modelling activities. These composite datasets may combine heterogeneous content such as legislation, news articles, and scientific publications. Each composite dataset includes its purpose, constituent data sources, collection rules, processing pipelines, and metadata characteristics. These descriptions complement the documentation of the individual constituent data sources.

- **Chapters 5-7, Individual Data Sources:** Documents individual data sources, organised into three categories based on their role and origin within the project. A data source is defined as a collection of data with a single primary data type (e.g., weather data, scientific publications, official monitoring results). The three data source categories are:
  - Project-Generated Data Sources (Chapter 5), comprising data collected, generated, curated, or transformed within EFRA that do not directly contribute to pilot workflows.
  - Decision-Support/Pilot-Contributing Data Sources (Chapter 6), including all data sources directly integrated into EFRA pilots and decision-support demonstrators, regardless of whether they originate from the project, industry, or external providers.
  - Collaborative/External Data Sources (Chapter 7), contributed by external projects or developed collaboratively with EFRA involvement, demonstrating cross-project synergies and reuse of established research assets.

To ensure consistency, comparability, and readability across all sections, each composite dataset and individual data source is documented using a unified structure composed of five elements:

- an Overview, describing the dataset, its relevance to the project, its size, and origin. Where applicable, datasets that participate in EFRA pilots include additional information describing their role within the decision-support workflows;
- Data Processing and Curation, outlining the collection rules, processing guidelines, applied ontologies or schemas;
- Metadata and Content, detailing the metadata fields, supported languages, and formats;
- Accessibility and Licensing, describing availability and usage conditions; and
- Contribution to Project KPIs, including a table summarising the dataset's contribution to the data related KPIs defined in section 3.4.
- **Chapter 8, Data Assets:** Summarizes all documented data sources, highlighting the evolution of their size over the course of the project and their contribution to the data related KPIs.
- **Chapter 9, Conclusions:** Summarizes the main findings regarding data coverage, diversity, integration, and their relevance to EFRA's KPIs.
- **Annex I:** Contains detailed tables presenting the values and calculations associated with KPI 1.1, KPI 1.2, and KPI 1.11 for each dataset.
- **Annex II:** Presents a snapshot of the EFRA central Excel-based data registry for documentation and monitoring across partners.

### 3. Data Population Methodology

#### 3.1. Overview and Scope

This chapter describes the methodology followed during Reporting Period 2 (RP2) for the systematic population, monitoring, and documentation of data sources within the EFRA ecosystem. The objective of this methodology was to ensure coordinated data acquisition across all EFRA partners, continuous tracking of progress, and transparent reporting of contributions towards the data related KPIs defined in the Grant Agreement.

To achieve this, EFRA adopted a structured data registry and monitoring process that enabled partners to document their datasets in a harmonised manner, while allowing the coordination team to continuously assess coverage, scale, and diversity of the populated data assets.

#### 3.2. Data Registry Tool for Data Population and Monitoring

A central Excel-based data registry was used throughout RP2 as the primary tool for documenting, monitoring, and coordinating data population activities across all EFRA partners. The registry was designed as a structured template capturing, for each dataset or data source, key descriptive metadata such as data type, origin, access conditions, update frequency, supported languages, and mapping to EFRA KPIs.

Importantly, the registry served not only as a reporting instrument, but as a living monitoring tool that was continuously updated by partners throughout the reporting period. Each partner was responsible for regularly updating their entries to reflect newly onboarded datasets, extensions of existing sources, or refinements in metadata and KPI contributions. A snapshot of the structure of the data registry, including its main fields and KPI mapping columns, is provided in Annex II.

#### 3.3. Monitoring, Coordination, and Progress Tracking During RP2

Data population progress was coordinated through a combination of recurring meetings and asynchronous communication. Partners were periodically reminded to update the registry, and the coordination team reviewed registry entries on a continuous basis to monitor progress and completeness.

This process enabled:

- Early identification of delays or missing information
- Clarification of dataset scope and categorisation
- Alignment across partners on KPI interpretation and reporting
- Timely decision-making when additional data acquisition actions were required to meet KPI targets

Offline communication via email complemented the meetings, allowing targeted follow-ups on specific datasets or KPI-related clarifications.

#### 3.4. KPI Monitoring Framework: Definitions and Scope

The Excel registry was explicitly structured to support continuous monitoring of all data related KPIs relevant to EFRA. Each dataset entry included fields mapping the dataset to one or more KPIs, as well as quantitative indicators (e.g., number of records, number of sources, supported languages) where applicable.

This enabled the consortium to compare the evolving contributions of individual datasets against the overall KPI targets defined in the Grant Agreement, and to assess cumulative progress throughout RP2.

The KPIs considered in this deliverable are defined as follows:

**KPI 1.1 - Data Types / Drivers Mined:** Covers the diversity of data types, spanning industrial, IT/IoT, scientific, business, administrative, environmental, and societal data. Examples include scientific publications, news articles, regulations, guidance, weather data, inspections, lab tests, prediction trends and risks.

**KPI 1.2 - Individual Multilingual Data Sources Mined Globally:** Counts the number of distinct data sources (organizations, authorities, platforms, feeds, repositories, etc.) ingested into the EFRA ecosystem, including multilingual sources from all world regions.

Examples: food safety authorities, scientific repositories, global monitoring systems, and additional inputs such as public media channels.

**KPI 1.3 - Aggregated Data Volume (Aggr. Data Vol.) (Including IoT Streams):** This KPI reflects the scale of data integrated within the EFRA ecosystem. It includes all harvested and partner-provided records across the Data Hub as well as the underlying datasets maintained by the partners' repositories. Time-series streams such as weather and sensor-based environmental measurements further contribute to the total dataset volume relevant for modelling contamination risks, crop conditions, and supply-chain exposure.

**KPI 1.4 - Social & News Data (Soc. & News Data):** Captures the volume of mined social media content and public online commentary relevant to food safety. Examples: customer reviews, online posts, and news articles.

**KPI 1.5 - Scientific & Expert Knowledge (Sci. & Expert Knowl.):** Counts data derived from scientific literature, domain-specific research outputs, and expert-curated reports. Examples: scientific publications, expert guidance documents.

**KPI 1.6 - Environmental & Climate Data (Env. & Clim. Data.):** Includes environmental data streams relevant to food safety, such as climatic variables, pollution indicators, weather station data, prediction trends and risks.

**KPI 1.7 - Economic & Trade Data (Econ. & Trade Data):** Covers economic-oriented datasets providing market, trade, and production insights. Examples: food product border rejections and import refusals, prediction trends, commodity prices, import/export data, ingredient-based trade flows, production records.

**KPI 1.8 - Regulatory & Legislative Data (Reg. & Legis. Data):** Counts data mined from official legal or regulatory sources, including laws, guidance, opinions, recalls, and border rejections. Examples: food regulations, guidance, food product recalls, inspections, food product border rejections and import refusals, foodborne disease outbreaks, predictions trends and risks.

**KPI 1.9 - Sustainability & Animal Health Data (Sust. & Animal Health Data):** Includes datasets related to sustainability, agricultural practices, pesticide/fertilizer use, and animal welfare. Examples: pest observations, fertilizer records, lab tests, inspections, prediction risks.

**KPI 1.11 - Languages Supported (Langs Supported):** Reports the languages supported by free-text content in each dataset, for example news articles available in English, German, and Greek.

### 3.5. Link to Per - Data Source Reporting

The monitoring methodology described above directly underpins the structure of the remainder of this deliverable. Each data source or composite dataset is described in a dedicated subsection, accompanied by a table summarising its contribution to the relevant KPIs. A consolidated overview of KPI values across all datasets is provided in

**Table 41** in section 8, enabling transparent aggregation and verification of EFRA's overall performance against data-related objectives, whereas Annex I documents the values of the KPI 1.1., KPI 1.2, and KPI 1.11.

## 4. Composite Datasets

This chapter presents composite datasets that integrate multiple types of data sources, such as legislation, news articles, and scientific publications, into a unified dataset. Although composite datasets are presented together in this chapter, they may fall under different data-source categories (e.g., Project-Generated, Decision-Support, or Collaborative) depending on how they are used within EFRA. For example, a composite dataset created for benchmarking NLP models is considered Project-Generated, while a composite dataset used directly in a pilot workflow is classified as Decision-Support / Pilot-Contributing.

### 4.1. SGS Digicomply Dataset

#### Overview

The SGS Digicomply dataset is a structured dataset that forms part of the Decision-Support / Pilot-Contributing Data Sources used in Use Case 3 (or UC#3), focusing on summaries and chatbot integration for regulatory documents. It has been employed for the manual validation of high-quality summaries, as well as for evaluating and comparing outputs from different LLMs to determine the most accurate summaries. Additionally, human validation was conducted on the text chunks that constitute the chatbot's responses, with multiple iterations performed in collaboration with domain experts.

The dataset is composed of roughly 124,000 records from 634 data sources containing English summaries of associated HTML documents, available both in their original language and in English, related to food safety, risk, and regulatory topics. The dataset includes four main data types: news, guidance, regulation, and scientific publications. Each HTML document has an associated summary that was either written by a human being or automatically generated using an AI tool for summarization.

#### Collection rules or processing pipelines

The SGS Digicomply dataset contains food-safety related documents collected by 24 May 2024 in the period M16-M20 of the project. The dataset spans the years 2002-2023, with approximately 71% in English, and the remaining in 60 other languages. The collected data and associated summaries were utilized to assess the outputs of various LLM models and to perform manual validation of the most relevant text segments for integration into the chatbot system.

The dataset is structured according to applied schemas in which each entry is annotated with the content topic (topic name), the content type (type name), and the source of the summary, indicating whether it was manually created or automatically extracted.

#### Metadata and Content

There are 60 languages supported by the dataset. The metadata fields are:

- Document type: scientific publications, governmental reports, news articles, regulatory publications.
- Topics: categories such as news, technical aspect, scientific part, laws and regulations, controls, enforcement actions, official notices and other thematic categories covering regulatory matters.
- Original title
- Content URLs
- Original content URL
- English content URL
- Type name
- Topic name
- Source name
- Summaries (both English and original) provided in three different ways:
  - Externally compiled summary (e.g., news summary or paper abstract).

- If no compiled summary is found, the first paragraph of the post is used.
- Internal SGS manual compiled summary for selected or reported posts (e.g., “top stories” relevant to the industry).

### **Dataset Types**

The dataset comprises the following data types:

- Food Safety News: 80,504
- Guidance: 20,346
- Food Regulations: 14,028
- Scientific Publications: 246

### **Access and Licensing**

The dataset is accessible through the EFRA Data Hub under the EFRA/SGS\_BENCHMARK/1 path.

### **Contribution to project KPIs**

Table 2 presents the contribution of the SGS Digicomply dataset to the data-related KPIs, showing how this dataset supports EFRA’s data-size and coverage targets.

**Table 2:** Contribution of the SGS Digicomply dataset to the data source related KPIs

KPI	Value	Description
1.1 Data Types	4	Data types
1.2 Sources	634	Data sources
1.3 Aggr. Data Vol.	127,204	Data records
1.4 Soc. & News Data	80,504	News articles
1.5 Sci. & Expert Knowl.	32,672	Guidance and scientific articles
1.8 Reg. & Legis. Data	46,454	Regulations and guidance
1.11 Langs Supp.	60	Natural languages

## **4.2. Chunk Relevance Validation Dataset**

### **Overview**

The Chunk Relevance Validation Dataset is a structured dataset to support the manual validation of the relevant chunks for the second scenario of the Use Case #3 generated in M26-M27. This is a subset of the main SGS Digicomply dataset, and it is composed of 1,330 records where the text is broken to several chunks, each of which manually evaluated for relevance.

### **Collection rules or processing pipelines**

The dataset comprises 1,330 records derived from 145 distinct data sources, each consisting of segmented content (chunks) extracted from the original documents. The dataset encompasses multiple information types, including news, guidance materials, regulatory documents, and scientific publications.

### **Metadata and Content**

The dataset is distributed in CSV format. Although the original documents span 60 languages, the dataset provides English text for all chunks. The metadata fields are:





- Query: segmented content
- Text: text in English
- Relevancy: manual annotation

### **Dataset Types**

The dataset comprises the following data types:

- Food Safety News: 96
- Guidance: 21
- Food Regulations: 25
- Scientific Publications: 3

### **Access and Licensing**

Original full documents are included internally for project experiments but are not publicly released.

### **Contribution to project KPIs**

**Table 3** presents the contribution of the Chunk Relevance Validation dataset to the data-source-related KPIs, highlighting how this dataset supports EFRA's monitoring of data-size progress and KPI target achievement.

**Table 3:** Contribution of the Chunk Relevance Validation dataset to the data source-related KPIs

KPI	Value	Description
1.1 Data Types	4	Data types
1.2 Sources	145	Data sources
1.3 Aggr. Data Vol.	1,330	Data records
1.4 Soc. & News Data	96	News articles
1.5 Sci. & Expert Knowl.	24	Guidance and scientific articles
1.8 Reg. & Legis. Data	25	Regulations and guidance
1.11 Langs Supp.	1	Natural languages

## 4.3. The FoodSafeSum Dataset

### **Overview**

The FoodSafeSum dataset is a structured composite dataset, added in M28 to the project, designed to support a wide range of Natural Language Processing (NLP) tasks that mirror real life food safety applications, including document categorization, evidence retrieval, event tracking, and question answering. It integrates expert-curated metadata, manual summaries, and Large Language Model (LLM)-generated summaries, enabling systematic analysis of food safety risks, regulatory content, and public communication. The dataset is owned and provided by SGS Digicomply and was created and released alongside the FoodSafeSum<sup>3</sup> (Enabling Natural Language Processing Applications for Food Safety Document Summarization and Analysis) paper, in which CNR, SU, and SGS Digicomply participated.

The dataset is taken as a subset from the SGS Digicomply dataset (see section 4.1) and is derived from 2,091 food safety-related documents collected by SGS Digicomply on 24 May 2024. These documents are sourced

<sup>3</sup> <https://aclanthology.org/2025.findings-emnlp.911/>



from news platforms, legal and regulatory agencies, food safety authorities and organizations, government guidance portals, scientific journals, and research platforms. The dataset spans the years 2002–2023, with approximately 65.7% in English, and the remainder in 33 other languages. Non-English documents were translated using Google Translate or DeepL and then manually curated for accuracy and relevance.

### **Data Processing and Curation**

Documents were curated as a subset from the SGS Digicomply dataset. Non-English documents were machine-translated and then manually reviewed for accuracy and relevance. LLM-generated summaries and titles were produced using the instruction-tuned LLaMA3-70B model (meta.llama3-70b-instruct-v1:0, via Amazon Bedrock). Controlled vocabularies are applied for hazard annotations, adapted from the Food Recall Incidents dataset (see **Section 5.1**).

### **Metadata and Content**

The dataset includes the following metadata and document versions:

- Metadata (from the SGS Digicomply dataset):
  - **Document Type:** News articles, regulations, governmental guidance, or scientific publications.
  - **Topics:** One or more of 12 thematic categories covering regulatory issues, scientific and technical aspects, consumer-facing content, and broader systemic concerns in food safety.
  - **Source name:** The name of the source the record was collected from.
- Automatically extracted hazard annotations, based on a controlled vocabulary adapted from the Food Recall Incidents dataset.
- Human-written or curated summary and title as taken from the SGS Digicomply dataset
- LLM-generated summary and title produced using the instruction-tuned LLaMA3-70B model (meta.llama3-70b-instruct-v1:0, via Amazon Bedrock).

The dataset is distributed in CSV format. Original full documents are included internally for project experiments but are not publicly released.

### **Dataset Types**

The dataset contains 2,091 documents with the following distribution across document types:

- News articles: 1,320
- Regulations: 352
- Governmental guidance: 288
- Scientific articles: 131

### **Access and Licensing**

The dataset is available on Zenodo under a CC BY 4.0 license (DOI: 10.5281/zenodo.17140845) and it is also accessible through the EFRA DataHub.

### **Contribution to project KPIs**

**Table 4** presents the contribution of the FoodSafeSum dataset to the data-source-related KPIs, illustrating how this composite dataset supports EFRA's data-size monitoring and KPI target assessment.

**Table 4:** Contribution of the FoodSafeSum dataset to the data source-related KPIs

KPI	Value	Description
1.1 Data Types	4	Data types
1.2 Sources	469	Data sources
1.3 Aggr. Data Vol.	2,091	Data records
1.4 Soc. & News Data	1,320	News articles
1.5 Sci. & Expert Knowl.	419	288 guidance + 131 scientific articles
1.8 Reg. & Legis. Data	640	352 regulations + 288 guidance
1.11 Langs Supp.	34	Natural languages

## 5. Project-Generated Data Sources

This chapter presents the data sources that were collected, generated, curated, or transformed within the EFRA project and that have not contributed directly to pilot workflows. These datasets were developed to support research, benchmarking, validation, and methodological experimentation across EFRA's scientific and technical activities. They include curated domain-specific datasets, systematic-review corpora, and subsets extracted from composite datasets such as FoodSafeSum. Together, they provide essential resources for NLP experimentation, classification tasks, ontology development, and cross-dataset comparisons, even though they are not integrated into the operational pilot scenarios.

### 5.1. Food Recall Incidents

#### Overview

Food Recall Incidents is a curated dataset of food recall announcements published by official food safety authorities. Each record pairs a short announcement title with the corresponding full recall text. The dataset is intended for text-mining and classification tasks focused on identifying the recalled product(s) and the hazard(s) motivating the recall. The dataset is owned by Agroknow and has been used within EFRA since milestone M15.

The dataset contains 7,546 title–text pairs in total:

- English: 6,644 records
- German: 888 records
- Other languages: French (8), Greek (4), Italian (1), Danish (1).

Text lengths range from 5–360 characters for titles and 56–48,318 characters for full texts. The data were collected via web crawling from 24 public food safety authority websites, mainly for recalls authored after 2010.

#### Data Processing and Curation

Data are collected via automated web crawling of official authority recall pages. Processing and enrichment include:

- manual expert annotation (1 expert + sampled verification by senior expert) of fine-grained product and hazard;
- coarse grained product-category and hazard-category derived using Agroknow-internal ontologies;
- automatic generation of hazard-title and product-title character-span features using feature importance from a Logistic Regression classifier. These supportive features indicate sample quality.
- Note on usability: fine-grained labels have low support for many classes, so additional preprocessing (e.g., clustering or filtering) may be required depending on use case.
- The schema consists of four linked classification targets:
  - hazard (261 fine-grained classes),
  - hazard-category (10 coarse classes),
  - product (1,256 fine-grained classes),
  - product-category (22 coarse classes).

Hazards and products are grouped into categories; the grouping is ontology-guided. For the product a custom vocabulary based on RASFF's terminology and linked with FoodEx2 is used, whereas for the hazard, again, a custom vocabulary based on RASFF's terminology and linked with the CAS one is used.

Provided metadata fields include release date (year, month, day), language, and country of issue. Quality assessment cues include dual-level label design (fine/coarse) and explicit warnings about class sparsity, supporting both robust modelling and transparency regarding label imbalance.

## Metadata and Content

The dataset is multilingual (6 languages), with English and German dominating. While multilingual training is possible, many downstream tasks in EFRA may focus on English due to coverage and label balance. The dataset is distributed as a single CSV file (food\_recall\_incidents.csv, ~17.7 MB). And the metadata fields are the following: title + full recall text + metadata + product/hazard labels + auxiliary span features (hazard-title, product-title).

## Accessibility and Licensing

- Public availability: deposited on Zenodo (DOI: 10.5281/zenodo.15007331).
- License: Creative Commons CC BY-NC-SA 4.0 (attribution, non-commercial use, share-alike).
- EFRA Data Hub: The dataset is available on the EFRA Data Hub.

## Contribution to project KPIs

**Table 5** presents the contribution of the Food Recall Incidents dataset to the data-source-related KPIs, showing how this dataset supports EFRA's data-size monitoring and KPI target assessment.

**Table 5:** Contribution of the Food Recall Incidents dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	24	Data sources
1.3 Aggr. Data Vol.	7,546	Data records
1.8 Reg. & Legis. Data	7,546	Scientific articles
1.11 Langs Supp.	6	Natural languages

## 5.2. Scientific Publications

### 5.2.1. Scientific Publications from FoodSafeSum Dataset

This data source is a part of the FoodSafeSum composite dataset (**Section 4.3**). This data source is derived from 131 scientific publications related to food safety, originating from 44 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in **Section 4.3**.

**Table 6** presents the contribution of the scientific publications from the FoodSafeSum dataset to the data-source-related KPIs, reflecting their role in EFRA's data-size tracking.

**Table 6:** Contribution of the scientific publications from FoodSafeSum dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	44	Data sources
1.3 Aggr. Data Vol.	131	Data records
1.8 Reg. & Legis. Data	131	Scientific articles
1.11 Langs Supp.	3	Natural languages

## 5.2.2. Abstracts of a Systematic Review on Food and Federated Learning

### **Overview**

The data source by WFSR is bibliographic records (titles, abstracts, keywords and associated metadata) of scientific publications retrieved from Google Scholar, Scopus, Web of Science, IEEE Xplore and CAB Abstracts. The dataset comprises abstracts, titles, and bibliometric information of all papers mentioning federated learning or federated approaches in combination with food- and agriculture-related terms, collected in two systematic search rounds and then screened on title and abstract by two independent researchers for relevance to food research use cases. Searches were run on July 31, 2023, and August 15, 2024, yielding 548 papers, of which 86 food-relevant papers were finally included; the dataset thus contains 548 screened abstracts (labelled relevant/not relevant) originating from those five bibliographic databases.

### **Data Processing and Curation**

Data processing consisted of: (1) structured query design combining “federated learning/federated approach” with extensive food-, safety-, and agriculture-related term blocks; (2) retrieval limited to title, abstract and keywords; (3) two-stage screening where two researchers independently judged relevance on title/abstract, with full-text scanning when relevance was unclear; and (4) inclusion of both journal and selected non-classic literature (conference papers, dissertations, preprints). Collection rules included time limitation to papers from 2016 onwards (first year of federated learning publications), inclusion only when a concrete food-domain use case or sufficiently detailed proposal was present, and exclusion of non-English papers and purely generic mentions; this implicitly defines the labelling criteria for “relevant to food” vs “not relevant”. The dataset is not mapped to any formal ontology or schema beyond standard bibliographic structures, nor does it present an explicit metadata completeness scoring pipeline; however, the systematic search strategy and explicit eligibility criteria constitute a basic quality assurance protocol for the labels.

### **Metadata and content**

For each publication, the available metadata fields comprise at least: database/source, search query block that retrieved it, title, abstract, keywords, publication year, publication type (journal article, conference, dissertation, preprint), and, for included papers, domain categorisation (e.g. food safety hazards, food quality, food production monitoring, food fraud, food supply chain, agriculture-oriented applications) and federated-learning framework attributes (server architecture, client type, data partitioning). The free-text content used for labelling is in English only, as non-English papers were excluded at screening. The database is presented as a CSV file.

### **Availability and licensing**

The dataset is publicly available on the EFRA Data Hub. The full text of the scientific articles is not available. But bibliometric information would allow them to obtain the full text from their publishers.

### **Contribution to project KPIs**

The dataset directly supports EFRA objectives by:

1. mapping the landscape of federated learning applications in food research across 86 implemented or proposed use cases;
2. identifying gaps such as the near absence of vertical federated learning, limited use of federated transfer learning, and lack of decentralized architectures; and
3. clarifying predominant drivers like data-sharing barriers and resource constraints that federated learning can mitigate.

**Table 7** presents the contribution of the systematic-review abstracts on food and federated learning to the data-source-related KPIs, highlighting how this dataset supports EFRA’s analytical objectives.

**Table 7:** Contribution of the abstracts of a systematic review on food and federated learning to the data source-related KPIs

KPI	Value	Description
1.2 Sources	5	Data sources
1.3 Aggr. Data Vol.	548	Data records
1.8 Reg. & Legis. Data	548	Scientific articles
1.11 Langs Supp.	1	Natural languages

### 5.2.3. Abstracts of a Systematic Review on Food and NLP

#### Overview

The dataset originates from abstracts of studies identified in a systematic review on NLP and food safety, collected via searches in CAB, IEEE, Scopus, Web of Science, and ACL Anthology on March 27, 2024. It describes 244 abstracts initially screened (after deduplication from 432 total results), with three annotators labelling them as relevant or not to the intersection of NLP and food safety. Created specifically for this survey, its size reflects the screening process leading to 120 full-text reviews and 95 final included studies. The data source is owned by WFSR.

#### Data Processing and Curation

The manual annotation was performed during the abstract screening: initial calibration on 50 random abstracts, followed by full labelling with discussion for disagreements. Collection rules followed PRISMA guidelines, prioritizing English studies with NLP pipelines focused on food safety; full-text annotations by five authors used a template capturing purpose, NLP task, data source, language, size, food matrices, and hazards. No ontologies or schemas applied; quality ensured via expert consensus and template standardization.

#### Metadata and Content

Metadata fields from the annotation template include binary relevance label, food safety purpose, NLP task, data source, language, number of data points, food matrices, and hazards. Free-text content is English abstracts from scientific publications; supported language is English per inclusion criteria. The dataset is presented as a CSV file.

#### Availability and Licensing

This dataset is presented as a standalone dataset on EFRA Data Hub. The title, abstract and bibliometric information of the paper is provided in the dataset.

#### Contribution to Project KPIs

Table 8 presents the contribution of the systematic-review abstracts on food and NLP to the data-source-related KPIs, illustrating their role in EFRA's data-size evaluation.

**Table 8:** Contribution of the abstracts of a systematic review on food and NLP to the data source related KPIs

KPI	Value	Description
1.2 Sources	5	Data sources
1.3 Aggr. Data Vol.	244	Data records
1.8 Reg. & Legis. Data	244	Scientific articles
1.11 Langs Supp.	1	Natural languages

## 5.3. News Articles

### 5.3.1. News Articles from FoodSafeSum Dataset

This data source is a part of the FoodSafeSum composite dataset (section 4.3). This data source is derived from 1,320 news articles related to food safety, originating from 353 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in section 4.3.

#### Contribution to Project KPIs

**Table 9** presents the contribution of the news articles from the FoodSafeSum dataset to the data-source-related KPIs, showing their impact on EFRA's data-size metrics.

**Table 9:** Contribution of the news articles from FoodSafeSum dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	469	Data sources
1.3 Aggr. Data Vol.	1,320	Data records
1.4 Soc. & News Data	1,320	News articles
1.11 Langs Supp.	20	Natural languages

### 5.3.2. News Articles by WFSR

#### Overview

New York Times (NYT) API was accessed to retrieve title, publication date, and leading sentences of around a million news articles. 2,000 of these articles were randomly selected and annotated as related to food or not. This annotation enables the development of the machine-learning classifiers for identifying food related text.

#### Data Processing and Curation

Each item in the dataset contains detailed meta-information retrieved from the NYT API. The title and leading sentences are used as the primary text fields for annotation and model development.

#### Metadata and Content

The dataset is in English. This dataset is prepared as a standalone CSV file. Metadata fields include headline, abstract, title, label, author, pub\_date, section\_name, lead, web\_url. The "label" field indicates whether the article was annotated as food-related or not.

#### Availability and Licensing

2,000 items and their annotations are shared on EFRA Data Hub.

#### Contribution to project KPIs

**Table 10** presents the contribution of the news articles by WFSR to the data-source-related KPIs, highlighting their role in supporting text-classification experimentation.

**Table 10:** Contribution of the news articles by WFSR to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data sources
1.3 Aggr. Data Vol.	2,000	Data records
1.4 Soc. & News Data	2,000	News articles
1.11 Langs Supp.	1	Natural languages

## 5.4. Governmental Guidance from FoodSafeSum Dataset

This data source is a part of the FoodSafeSum composite dataset (section 4.3). This data source is derived from 288 governmental guidance documents related to food safety, originating from 89 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in section 4.3.

### Contribution to project KPIs

**Table 11** presents the contribution of the governmental guidance from the FoodSafeSum dataset to the data-source-related KPIs, showing how this subset contributes to EFRA's data-size monitoring.

**Table 11:** Contribution of the governmental guidance from FoodSafeSum dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	89	Unique data sources
1.3 Aggr. Data Vol.	288	Data records
1.5 Sci. & Expert Knowl.	288	Guidance
1.8 Reg. & Legis. Data	288	Guidance
1.11 Langs Supp.	8	Natural languages

## 5.5. Regulations from FoodSafeSum Dataset

This data source is a part of the FoodSafeSum composite dataset (section 4.3). This data source is derived from 352 regulations related to food safety, originating from 98 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in section 4.3.

### Contribution to project KPIs

**Table 12** presents the contribution of the regulations from the FoodSafeSum dataset to the data-source-related KPIs, reflecting their role in EFRA's regulatory-content coverage.

**Table 12:** Contribution of the regulations from FoodSafeSum dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	98	Unique data sources
1.3 Aggr. Data Vol.	352	Data records
1.8 Reg. & Legis. Data	352	Regulations
1.11 Langs Supp.	17	Natural languages



## 6. Decision-Support / Pilot-Contributing Data Sources

This chapter presents the data sources collected, generated, curated, or transformed within the EFRA project that support the project's decision-support use cases and pilots. The chapter is structured around EFRA's core use cases, with each subsection describing the datasets relevant to a specific use case.

### 6.1. Use Case 1 - Data Sources for the EFRA Decision Support Tool

#### 6.1.1. Food Product Recalls

##### Overview

Food product recalls contain data on food unfit for human or animal consumption announced at national level following food safety testing performed on products directly available to consumers. These data are announced by food safety authorities worldwide and, following post-processing, are made available in FOODAKAI and Agroknow's data API. The dataset contains 126,907 product recalls and is owned by Agroknow.

##### Data Processing and Curation

The data are collected in a fully automated way by crawling the websites in which they are announced. Following the collection step, they undergo a set of automatic enrichment steps to identify the core metadata. The results of the automated enrichment steps are reviewed by human experts (food scientists) before being made available in FOODAKAI.

The core metadata properties and respective vocabularies/ontologies used are the following:

- Date of announcement,
- Ingredient that caused the recall. A custom vocabulary based on RASFF's terminology and linked with FoodEx2 is used,
- Hazard, reason behind the recall. Again, a custom vocabulary based on RASFF's terminology and linked with the CAS one is used,
- Country of origin of the recalled ingredient,
- Company involved in the recall, when available.

##### Metadata and Content

The textual content is available in 16 languages. The core metadata fields are mentioned above.

##### Availability and Licensing

Data are made available in FOODAKAI and Agroknow's data API.

##### Contribution to project KPIs

**Table 13** presents the contribution of the food product recalls to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 13:** Contribution of the food product recalls to the data source related KPIs

KPI	Value	Description
1.2 Sources	34	Data sources
1.3 Aggr. Data Vol.	126,907	Data records

1.8 Reg. & Legis. Data	126,907	Product Recalls
1.11 Langs Supp.	16	Natural languages

### 6.1.2. Global Food Product Border Rejections

#### **Overview**

Global food product border rejections contain data on food products unfit for human or animal consumption that result from testing carried out at the customs of each country for food products imported from other countries. This dataset is very similar to food product recalls (section 6.1.1) and includes 598,738 data records and is owned by Agroknow.

#### **Data Processing and Curation**

The way of collection and review is the same as for food product recalls. The core metadata and vocabularies employed are the same as those used for food product recalls.

#### **Metadata and Content**

The metadata fields are the same as those used for food product recalls. The border rejection records may be available in one of four languages.

#### **Availability and Licensing**

Data are made available in FOODAKAI and Agroknow's data API.

#### **Contribution to project KPIs**

**Table 14** presents the contribution of global food product border rejections to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 14:** Contribution of the global food product border rejections to the data source related KPIs

KPI	Value	Description
1.2 Sources	10	Data sources
1.3 Aggr. Data Vol.	598,738	Data records
1.7 Econ. & Trade Data	598,738	Product border rejections
1.8 Reg. & Legis. Data	598,738	Product border rejections

### 6.1.3. Food Safety News

#### **Overview**

Food safety news dataset consists of news items published by news websites worldwide, covering topics such as product recalls and developments in food safety. The dataset is owned by Agroknow. The data are collected and processed in a fully automated manner, without a human validation step. The dataset is collected from Food Safety News, Food Safety Tech, and MEDISYS and contains 94,953 news articles.

#### **Data Processing and Curation**

The data are collected and processed in a fully automated way, without the involvement of a human validation step. There is no specific ontology or schema used for the metadata.

### **Metadata and Content**

The textual content is available in English. The core metadata fields included in the dataset are:

- Date of news post
- Title
- Description / body

### **Availability and Licensing**

The data are made available in FOODAKAI and Agroknow's data API.

### **Contribution to project KPIs**

**Table 15** presents the contribution of the food safety news to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 15:** Contribution of the food safety news to the data source related KPIs

KPI	Value	Description
1.2 Sources	3	Data sources
1.3 Aggr. Data Vol.	95,076	Data records
1.4 Soc. & News Data	95,076	News articles

## 6.1.4. Foodborne Disease Outbreaks

### **Overview**

Foodborne Disease Outbreaks contain information on food unfit for human consumption that led to multiple people becoming sick, hospitalized, or potentially dying. The dataset is owned by Agroknow. The dataset is collected from U.S. Food & Drug Administration (FDA), and CDC and contains 3,889 outbreak records.

### **Data Processing and Curation**

The collection and review process is identical to that used for food product recalls and global food product border rejections. The metadata properties and vocabularies are shared with food product recalls and border rejections, with the following extensions specific to outbreak data:

- Number of cases
- Number of hospitalizations
- Number of deaths

### **Availability and Licensing**

The data are made available through FOODAKAI and Agroknow's data API.

### **Contribution to project KPIs**

**Table 16** presents the contribution of the foodborne disease outbreaks to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 16:** Contribution of the foodborne disease outbreaks to the data source related KPIs

KPI	Value	Description
1.2 Sources	2	Data sources
1.3 Aggr. Data Vol.	4,064	Data records
1.8 Reg. & Legis. Data	4,064	Foodborne disease outbreaks

### 6.1.5. Inspections

#### Overview

The Inspections dataset contains information on inspections performed by Food Safety Authorities at the premises or plant locations of food companies, along with the specific findings. The dataset is by Agroknow. The dataset is collected from the U.S. Food & Drug Administration (FDA) and contains 501,949 inspection records.

#### Data Processing and Curation

The data are collected and made available in FOODAKAI and Agroknow's data API in a fully automated way, without a human review step. No specific vocabularies or ontologies are used in this dataset.

#### Metadata and Content

The core metadata fields of inspections are:

- Date of inspection
- Country of the inspected company
- Company that underwent the inspection
- Inspection outcome

#### Availability and Licensing

The data are made available through FOODAKAI and Agroknow's data API.

#### Contribution to project KPIs

**Table 17** presents the contribution of the inspections to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 17:** Contribution of the inspections to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data sources
1.3 Aggr. Data Vol.	501,949	Data records
1.8 Reg. & Legis. Data	501,949	Inspections
1.9 Sust. & Animal Health Data	501,949	Inspections

### 6.1.6. Lab Tests

#### Overview

Lab tests contain results from pesticide or biological monitoring programs conducted by Food Safety Authorities worldwide. The dataset is owned by Agroknow. The dataset is collected from EFSA and contains 226,792,288 lab test records.

#### Data Processing and Curation

The collection and processing of lab tests are fully automated, with predefined mappings applied between raw data and the target schema during processing. Lab tests share the same metadata properties and vocabularies as food product recalls and border rejections.

#### Metadata and Content

The core metadata fields for lab tests are:

- Lab test numerical result
- LOD (limit of detection value)
- LOQ (limit of qualification value)
- Lab test result classification (above/below limit or unquantified)

#### Availability and Licensing

The data are made available through FOODAKAI and Agroknow's data API.

#### Contribution to project KPIs

**Table 18** presents the contribution of the lab tests to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 18:** Contribution of the lab tests to the data source related KPIs

KPI	Value	Description
1.2 Sources	37	Data sources
1.3 Aggr. Data Vol.	226,792,288	Data records
1.6 Env. & Clim. Data.	226,792,288	Lab tests
1.9 Sust. & Animal Health Data	226,792,288	Lab tests

### 6.1.7. Weather Data

#### Overview

Weather data contain information on weather conditions in countries worldwide on a daily basis. The dataset is owned and provided by Agroknow. The data are collected automatically by an external API service (Visual Crossing) and made available through Agroknow's API. They are also utilized by prediction models to generate the data described in subsequent sections. The dataset is collected from EFSA and contains 1,925,339 records.

#### Data Processing and Curation

The data are collected automatically via an external API service. No ontologies or vocabularies are used for the weather data properties.

### **Metadata and Content**

The dataset does not contain free text. The core metadata fields for Weather Data are:

- Date of weather conditions reported
- Country
- Precipitation
- Average, minimum, and maximum temperature
- Humidity
- Windspeed
- Cloudcover

### **Availability and Licensing**

The data are made available through Agroknow's data API.

### **Contribution to project KPIs**

**Table 19** presents the contribution of the weather data to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 19:** Contribution of the weather data to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data sources
1.3 Aggr. Data Vol.	1,925,339	Data records
1.6 Env. & Clim. Data.	1,925,339	Weather data

## 6.1.8. Prediction Trends

### **Overview**

Prediction trends include the results of Agroknow's predictive analytics models, showing the expected number of incidents (food recalls and border rejections) for any combination of ingredient, hazard, and country of origin. The dataset is owned by Agroknow. These data are generated automatically, with evaluation performed using standard data science approaches such as cross-validation, precision, and recall. The dataset is generated by predictive models, and it contains 42,992,662 records.

### **Data Processing and Curation**

The data are generated automatically by predictive analytics models. The core metadata properties and vocabularies are:

- Date of prediction result (future date)
- Ingredient / food product the prediction is about (custom vocabulary based on RASFF terminology, linked with FoodEx2)
- Hazard the prediction is about (custom vocabulary based on RASFF terminology, linked with CAS)
- Country the prediction is about
- Number of expected incidents
- 3-month trend value
- 6-month trend value
- 12-month trend value

### **Metadata and Content**

The core metadata fields are the same as mentioned above. The dataset does not contain free text.

### **Availability and Licensing**

The data are made available through Agroknow's API.

### **Contribution to project KPIs**

**Table 20** presents the contribution of the prediction trends to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 20:** Contribution of the prediction trends to the data source related KPIs

KPI	Value	Description
1.3 Aggr. Data Vol.	42,992,662	Data records
1.6 Env. & Clim. Data.	42,992,662	Prediction trends
1.7 Econ. & Trade Data	42,992,662	Prediction trends
1.8 Reg. & Legis. Data	42,992,662	Prediction trends

## 6.1.9. Prediction Risks

### **Overview**

Prediction risks incorporate the results of prediction trends together with FOODAKAI's risk assessment formula (hazard severity × probability) to produce the expected food safety risk assessment value for ingredient and hazard combinations. The dataset is owned by Agroknow. The data are produced automatically by predictive models without a human review or validation step. The dataset contains 3,052,452 records.

### **Data Processing and Curation**

The data are generated automatically by predictive models. The core metadata properties are:

- Date of the risk assessment result (future date)
- Ingredient the risk assessment result is about
- Hazard the risk assessment result is about

### **Metadata and Content**

The core metadata fields are the same as mentioned above. The dataset does not contain free text.

### **Availability and Licensing**

The data are made available through Agroknow's data API.

### **Contribution to project KPIs**

**Table 21** presents the contribution of the prediction risks to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 21:** Contribution of the prediction risks to the data source related KPIs

KPI	Value	Description
1.3 Aggr. Data Vol.	3,052,452	Data records
1.6 Env. & Clim. Data.	3,052,452	Prediction risks
1.8 Reg. & Legis. Data	3,052,452	Prediction risks
1.9 Sust. & Animal Health Data	3,052,452	Prediction risks

### 6.1.10. Production Data

#### Overview

Data about the production volumes & statistics officially announced by FAO. The data is processed and enriched with terms from product and country taxonomies.

#### Data Processing and Curation

The collection and processing of Production Data is fully automated, with predefined mappings applied between raw data and the target schema during processing. Production data have a specific set of metadata properties and vocabularies, and they share some core metadata properties with food product recalls and border rejections.

#### Metadata and Content

The core metadata fields for production data are:

- Food product
- Country
- Date
- Volume
- Data source

#### Availability and Licensing

The data are made available through FOODAKAI and Agroknow's data API.

#### Contribution to project KPIs

**Table 22** presents the contribution of the production data to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 22:** Contribution of the production data to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data source
1.3 Aggr. Data Vol.	503,329	Data records
1.7 Econ. & Trade Data	503,329	Production data



### 6.1.11. Trade Data

#### **Overview**

Data about trade data for all agricultural commodities and products officially announced by FAO. The data is processed and enriched with terms from product and country taxonomies.

#### **Data Processing and Curation**

The collection and processing of Trade Data is fully automated, with predefined mappings applied between raw data and the target schema during processing. Production data have a specific set of metadata properties and vocabularies, and they share some core metadata properties with food product recalls and border rejections.

#### **Metadata and Content**

The core metadata fields for trade data are:

- Food product
- Country of export
- Country of import
- Date
- Volume
- Data Source

#### **Availability and Licensing**

The data are made available through FOODAKAI and Agroknow's data API.

#### **Contribution to project KPIs**

**Table 23** presents the contribution of the trade data to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 23:** Contribution of the trade data to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data source
1.3 Aggr. Data Vol.	41,424,465	Data records
1.7 Econ. & Trade Data	41,424,465	Trade data

### 6.1.12. Price Data

#### **Overview**

Aggregated price data officially announced by international sources and bodies that monitor agricultural products prices.

#### **Data Processing and Curation**

The collection and processing of trade data is fully automated, with predefined mappings applied between raw data and the target schema during processing. Production data have a specific set of metadata properties and vocabularies and they share some core metadata properties with food product recalls and border rejections.

### **Metadata and Content**

The core metadata fields for trade data are:

- Food product
- Country
- Price
- Date
- Data Source

### **Availability and Licensing**

The data are made available through FOODAKAI and Agroknow's data API.

### **Contribution to project KPIs**

**Table 24** presents the contribution of the price data to the data-source-related KPIs, showing how this dataset supports EFRA's data-size monitoring and KPI target assessment.

**Table 24:** Contribution of the price data to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data source
1.3 Aggr. Data Vol.	415,670	Data records
1.7 Econ. & Trade Data	415,670	Price data

## 6.1.13. Poultry Production Data

### **Overview**

Moy Park provided 2 proprietary poultry production datasets collected directly from their integrated farming operations. Dataset 1 contains 10,376 records covering the years 2022–2023, while Dataset 2 includes 39,714 records spanning 2020–2025. Each record corresponds to a flock housed within a specific poultry house and captures detailed information on production practices, growth performance, health indicators, and laboratory testing results.

### **Data Processing and Curation**

The datasets were ingested and curated through a controlled, semi-automated workflow. Predefined mappings aligned the raw Excel fields with the target schema, ensuring consistency across farm identifiers, temporal attributes, health indicators, and performance metrics. Additional manual validation was applied to harmonise date formats, resolve inconsistencies in flock identifiers, and ensure the correct interpretation of production and laboratory variables.

These data represent the sole farm-level dataset available within EFRA, offering comprehensive production, growth, and health indicators essential for the poultry demonstrator.

### **Metadata and Content**

The datasets include a comprehensive set of production, health, and laboratory fields. The core metadata and content categories include:

- Farm and Flock Identification (e.g., FiscalTag, Growout No, Farm No, Entity No, House No, Complex Entity Number)
- Breed and Hatchery Information (Breed No, Hatchery No, Product No)

- Placement and Growth Data (First Date Placed, Head Placed, 7/14/21/28 Day Weights, Thin/Clear Dates and Weights)
- Health and Welfare Indicators (Hock %, Podo %, HealthStatus, SerotypeNo)
- Mortality and Culling Metrics (D3/D7/D14/D21/D28/LOF Cull %, Mort %, Total ACM %)
- Laboratory and Sampling Data (SwabDate, Sample Receive Date, Swab Date Diff)
- Production and Processing Information (WITHDRAWAL, Grower, Finisher, CRUMB, Kill Plant, Ave Weight, Ave Age)

These fields collectively provide a detailed view of flock performance, health evolution, and potential contamination events across the production cycle.

### **Availability and Licensing**

The Moy Park datasets are private, confidential, and not publicly accessible. They were shared exclusively for EFRA research purposes and used to develop the poultry demonstrator, including the weight-gain and mortality-rate classification models.

### **Contribution to project KPIs**

**Table 25** summarises the contribution of the Moy Park poultry production data to the data-source-related KPIs, particularly in relation to model development, demonstrator validation, and the integration of industry-grade farm-level data.

**Table 25:** Contribution of the Moy Park poultry production data to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data source
1.3 Aggr. Data Vol.	50,090	Data records
1.9 Sust. & Animal Health Data	50,090	Poultry data

## **6.1.14. Food Fortress Dataset**

### **Overview**

The Food Fortress dataset consists of laboratory analyses of mycotoxins in animal feed collected through the Food Fortress quality assurance scheme in Northern Ireland. The scheme is an industry-wide collaboration designed to enhance feed and food chain safety through strategic sampling and testing of imported materials and finished feeds. It is recognised internationally as a world-leading risk-management and feed-quality assurance programme supported by the Institute for Global Food Security at Queen's University BelfastQueen's University.

For EFRA, Food Fortress provided 864 private records covering mycotoxin concentrations in compound feeds formulated for poultry, pigs, beef cattle, and ruminants. These data represent a subset of the broader Food Fortress monitoring programme, which covers close to 5 million tonnes of compound feed production annually across more than 80 participating companies.

### **Data Processing and Curation**

The dataset was ingested and curated using a semi-automated workflow. Predefined mappings aligned the raw Excel fields with the target schema, ensuring consistent representation of mycotoxin categories, concentration values, sampling metadata, and regulatory thresholds. Manual validation steps ensured correct interpretation of units ( $\mu\text{g/kg}$ ), harmonisation of date formats, and verification of sample identifiers.

These data constitute the laboratory-based contaminant dataset within EFRA and provide essential quantitative evidence for the development of the mycotoxin decision support tool in Use Case 1.

### **Metadata and Content**

The dataset includes detailed analytical and contextual information for each sample. The core metadata fields include:

- Mycotoxin Category (e.g., DON, Zearalenone, Aflatoxin B1, Ochratoxin A)
- Date
- Composite\_Size
- Sample\_Type
- Sample\_Ref
- Issue Date
- Result (µg/kg)
- Internal Limit (µg/kg)
- Legal Limit (µg/kg)
- Guidance Limit (µg/kg)
- file\_path

### **Availability and Licensing**

The Food Fortress dataset is private and confidential, shared exclusively for EFRA research purposes. It originates from the Food Fortress network; a collaboration of feed producers and importers committed to improving feed safety through strategic sampling and testing

### **Contribution to project KPIs**

**Table 26** summarises the contribution of the food fortress mycotoxin data to the data-source-related KPIs, particularly in relation to contaminant monitoring, risk assessment, and the development of the mycotoxin decision support tool.

**Table 26:** Contribution of the food fortress dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data source
1.3 Aggr. Data Vol.		Data records
1.9 Sust. & Animal Health Data		Lab tests

## 6.1.15. Global Fraud Reports

### **Overview**

Global fraud reports contain data on fraud cases on food products and ingredients at a national level. These reports are the results of announcements of food safety authorities as well as news pieces made available worldwide by news websites and are made available in FOODAKAI and Agroknow's data API. The dataset contains 147,363 data records and is owned by Agroknow.

### **Data Processing and Curation**

The data are collected in a fully automated way by crawling the websites in which they are announced and publicly available news aggregators. Following the collection step, they undergo a set of automatic

enrichment steps to identify the core metadata. The results of the automated enrichment steps are reviewed by human experts (food scientists) before being made available in FOODAKAI.

The core metadata properties and respective vocabularies/ontologies used are the following:

- Date of announcement,
- Ingredient involved in the fraud case. A custom vocabulary based on RASFF's terminology and linked with FoodEx2 is used,
- Hazard, fraud reason behind the recall. Again, a custom vocabulary based on RASFF's terminology and linked with the CAS one is used,
- Country of origin of the fraudulent ingredient,
- Company involved in the case, when available.

### **Metadata and Content**

The textual content is available in 162 languages. The core metadata fields are mentioned above.

### **Availability and Licensing**

Data are made available in FOODAKAI and Agroknow's data API.

### **Contribution to project KPIs**

**Table 27** presents the contribution of the fraud reports to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 27:** Contribution of the fraud reports to the data source related KPIs

KPI	Value	Description
KPI 1.2 Sources	38	Data sources
KPI 1.3 Aggr. Data Vol.	147,363	Data entries
KPI 1.4 Soc. & News Data	1,887	Fraud cases on news items
KPI 1.8 Reg. & Legis. Data	145,476	Fraud cases by food safety authorities

## **6.2. Use Case 2 - Data Sources for the EFRA Decision Support Tool**

### **6.2.1. Weather Data**

#### **Overview**

The Weather Data dataset by Agrivi includes historical weather information covering the past three years up to the date of export (December 2025). The data are collected via an internally developed weather reporting application within the AGRIVI platform, which connects to an internal weather service backed by a third-party weather data provider integrated into AGRIVI software in order to retrieve field-specific weather parameters. The dataset comprises over two million data points covering more than 600 field locations in total and is used within the project to support predictive modelling for pest risk and other decision-support workflows. Part of these data were used to train machine learning models.

#### **Data Processing and Curation**

Weather data are collected using an internal .NET Core 3.1 application, which retrieves field metadata (Field ID, Company ID, field name, geographic coordinates) and then queries an internal weather service for

historical weather parameters. Temporary JSON files store retrieved data before consolidation into a single CSV file, where each Field ID is stored in a separate worksheet.

### **Metadata and Content**

The dataset does not contain free text and is provided in CSV format. The following weather parameters are collected:

- DateMeasured
- WindSpeedAvg, WindSpeedMax, WindSpeedMin
- DewPointAvg
- RelativeHumidityAvg, RelativeHumidityMax, RelativeHumidityMin
- PrecipitationAccumulatedPeriod
- AirTempAvg, AirTempMax, AirTempMin

### **Access and Licensing**

The dataset is accessible through the EFRA DataHub under the EFRA/AgriviWeatherReport/1 path.

### **Contribution to project KPIs**

Table 28 presents the contribution of the weather data to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 28:** Contribution of the weather data by Agrivi to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data sources
1.3 Aggr. Data Vol.	2,000,000	Data records
1.6 Env. & Clim. Data.	2,000,000	Weather data

## 6.2.2. Pest Observations

### **Overview**

This data source relates to weather-driven inputs used to support pest observation and prediction workflows within the AGRIVI platform. The data are provided by a third-party satellite-based service, DNT (Data Network for Technologies), which delivers high-resolution, field-specific weather information, including both current and forecast data. These data are integrated into the AGRIVI platform to support pest risk modelling and decision-support functionalities. Weather data integration for selected pilot fields started at the beginning of the second piloting cycle and has since been updated on a daily basis, with daily forecast and environmental condition data being actively used from May, 2025 onwards. The dataset originates from the DNT satellite-based weather service and covers all pilot fields and regions involved in the AGRIDRONE project across Croatia, Romania, and Bulgaria. It is provided at a daily, per-field resolution and comprises dozens of weather data points collected per field per day over the duration of the pilot.

The weather data is a core input for AGRIVI's pest prediction algorithm. Environmental variables such as temperature, humidity, precipitation, and wind speed act as triggers for pest alarm activation. When specific weather conditions known to support pest development are detected, the system generates pest alerts. These alerts are visible to farmers and agronomists, prompting timely scouting activities, proactive pest management, and a reduction in pesticide use, since interventions are targeted rather than broad-spectrum. This process improves both the efficiency and precision of pest detection and control.

The dataset is fully integrated into the EFRA pilots. Data are mapped to specific field coordinates, and the system processes the information daily to update the pest alarm engine, which runs in the background and evaluates risk levels based on predefined thresholds. No additional user interaction is required, as weather data are seamlessly incorporated to enhance decision support.

The integration of weather data has directly supported updates to the pest alarm system, including algorithm refinement for the following high-impact pests:

#### Corn Pests:

- Blue cereal leaf beetle
- European corn borer
- Fusarium ear rot
- Fusarium stalk rot
- Northern corn leaf blight
- Red cereal beetle

#### Wheat Pests:

- Red cereal leaf beetle (*Oulema melanopus* L.)
- Blue cereal leaf beetle (*Oulema lichenis* L.)

#### Sunflower Pest:

- Black bean aphid (*Aphis fabae* Scop.)

These updates were driven by field observations during the pilot, matched against weather patterns. This significantly improved the relevance and accuracy of pest alerts in the 2025 cycle.

### **Data Processing and Curation**

Weather data are transmitted automatically on a daily basis from the DNT service and ingested into the AGRIVI platform. Upon ingestion, the data are matched to specific field locations using GPS coordinates. Internal thresholds and pest appearance criteria are then applied to process the incoming weather measurements into risk-level indicators and alerts. Pest-specific condition sets are defined based on established agronomic standards and are further refined using observations and evidence collected during the pilot activities.

### **Metadata and Content**

The dataset does not contain any free-text fields. Each weather data record consists of a timestamp, the corresponding field location, the measurement type (such as temperature or humidity), and the associated value. The data are internally processed through an API in JSON format and stored in structured database fields within AGRIVI's cloud-based system.

### **Availability and Licensing**

The weather data are obtained under a commercial license from DNT and are therefore not open access. The dataset itself is not made available through the EFRA Data Hub or the EFRA Marketplace. The data are stored within AGRIVI's cloud-based infrastructure in structured form and accessed internally via APIs. While the raw weather data cannot be shared, aggregated insights and results derived from these data are made available to EFRA partners and are included in project deliverables and pilot reports.

### **Contribution to project KPIs**

**Table 29** presents the contribution of the pest observations to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.



**Table 29:** Contribution of the pest observations to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data sources
1.3 Aggr. Data Vol.	1,613	Data records
1.6 Env. & Clim. Data.	1,613	Pest observations
1.9 Sust. & Animal Health Data	1,613	Pest observations

## 6.3. Use Case 3 - Data Sources for the EFRA Decision Support Tool

### 6.3.1. Scientific Publications

#### 6.3.1.1. Scientific Publications from SGS Digicomply Dataset

This data source is a part of the SGS Digicomply composite dataset (**Section 4.1**) used in Use Case #3, focusing on summaries and chatbot integration for regulatory documents. This data source is derived from 246 scientific publications related to food safety, originating from 26 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in section 4.1.

#### Contribution to project KPIs

**Table 30** presents the contribution of the scientific publications from SGS Digicomply dataset to the data-source related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 30:** Contribution of the scientific publications from SGS Digicomply Dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	26	Data sources
1.3 Aggr. Data Vol.	246	Data records
1.5 Sci. & Expert Knowl.	246	Scientific articles

#### 6.3.1.2. Scientific Publications from Chunk Relevance Validation Dataset

This data source is a part of the Chunk Relevance Validation composite dataset (section 4.2). It supports the manual validation of the relevant chunks for the second scenario of the Use Case #3, generated in M26-M27. This data source is derived from 3 scientific publications related to food safety. All metadata, annotations, and summaries associated with this data source are presented in section 4.2.

#### Contribution to project KPIs

**Table 31** presents the contribution of the scientific publications of the Chunk Relevance Validation dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 31:** Contribution of the Chunk Relevance Validation dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	3	Data sources
1.3 Aggr. Data Vol.	3	Data records
1.5 Sci. & Expert Knowl.	3	Scientific articles

## 6.3.2. News Articles

### 6.3.2.1. News Articles from SGS Digicomply Dataset

This data source is a part of the SGS Digicomply composite dataset (section 4.1.) used in Use Case #3, focusing on summaries and chatbot integration for regulatory documents. This data source is derived from 80,504 news articles related to food safety, originating from 452 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in section 4.1.

#### Contribution to project KPIs

**Table 32** presents the contribution of the news articles from SGS Digicomply dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 32:** Contribution of the news articles from SGS Digicomply dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	452	Data sources
1.3 Aggr. Data Vol.	80,504	Data records
1.4 Soc. & News Data	80,504	News articles

### 6.3.2.2. News Articles from Chunk Relevance Validation Dataset

This data source is a part of the Chunk Relevance Validation composite dataset (section 4.2). It supports the manual validation of the relevant chunks for the second scenario of the Use Case #3, generated in M26-M27. This data source is derived from 96 news articles related to food safety. All metadata, annotations, and summaries associated with this data source are presented in section 4.2.

#### Contribution to project KPIs

**Table 33** presents the contribution of the news articles of the Chunk Relevance Validation dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 33:** Contribution of the news articles of the Chunk Relevance Validation dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	3	Data sources
1.3 Aggr. Data Vol.	3	Data records
1.5 Sci. & Expert Knowl.	3	Scientific articles

## 6.3.3. Governmental Guidance

### 6.3.3.1. Governmental Guidance from SGS Digicomply Dataset

This data source is a part of the SGS Digicomply composite dataset (section 4.1.) used in Use Case #3, focusing on summaries and chatbot integration for regulatory documents. This data source is derived from 20,346 governmental guidance documents related to food safety, originating from 65 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in Section 4.1.

### **Contribution to project KPIs**

**Table 34** presents the contribution of the governmental guidance from SGS Digicomply dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 34:** Contribution of the governmental guidance from SGS Digicomply dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	65	Data sources
1.3 Aggr. Data Vol.	20,346	Data records
1.5 Sci. & Expert Knowl.	20,346	Guidance
1.8 Reg. & Legis. Data	20,346	Guidance

#### 6.3.3.2. Governmental Guidance from Chunk Relevance Validation Dataset

This data source is a part of the Chunk Relevance Validation composite dataset (section 4.2). It supports the manual validation of the relevant chunks for the second scenario of the Use Case #3, generated in M26-M27. This data source is derived from 96 news articles related to food safety. All metadata, annotations, and summaries associated with this data source are presented in section 4.2.

### **Contribution to project KPIs**

**Table 35** presents the contribution of the governmental guidance from Chunk Relevance Validation dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 35:** Contribution of the governmental guidance from Chunk Relevance Validation dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	21	Data sources
1.3 Aggr. Data Vol.	21	Data records
1.5 Sci. & Expert Knowl.	21	Guidance
1.8 Reg. & Legis. Data	21	Guidance

#### 6.3.4. Regulations

##### 6.3.4.1. Regulations from SGS Digicomply Dataset

This data source is a part of the SGS Digicomply composite dataset (section 4.1.) used in Use Case #3, focusing on summaries and chatbot integration for regulatory documents. This data source is derived from 14,028 regulations related to food safety, originating from 91 distinct sources. All metadata, annotations, and summaries associated with this data source are presented in section 4.1.

### **Contribution to project KPIs**

**Table 36** presents the contribution of the regulations from SGS Digicomply dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 36:** Contribution of the regulations from SGS Digicomply dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	91	Data sources
1.3 Aggr. Data Vol.	14,028	Data records
1.8 Reg. & Legis. Data	14,028	Regulations
1.2 Sources	91	Data sources

#### 6.3.4.2. Regulations from Chunk Relevance Validation Dataset

This data source is a part of the Chunk Relevance Validation composite dataset (section 4.2). It supports the manual validation of the relevant chunks for the second scenario of the Use Case #3, generated in M26-M27. This data source is derived from 25 regulations related to food safety. All metadata, annotations, and summaries associated with this data source are presented in section 4.2.

#### Contribution to project KPIs

**Table 37** presents the contribution of the regulations from Chunk Relevance Validation dataset to the data-source-related KPIs, showing how this data source supports EFRA's data-size monitoring and KPI target assessment.

**Table 37:** Contribution of the regulations from Chunk Relevance Validation dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	25	Data sources
1.3 Aggr. Data Vol.	25	Data records
1.8 Reg. & Legis. Data	25	Regulations
1.2 Sources	25	Data sources

## 7. Collaborative / External Data Sources

This chapter presents data sources contributed by external projects as well as data sources created collaboratively with EFRA involvement and partner projects. These data sources demonstrate cross-project synergy and the efforts to advance research in the food safety field, enriching the project's data sources.

### 7.1. Contextualized Social Recommender System Data Source (CoSRec)

#### **Overview**

The CoSRec dataset, developed by CNR, is a collection of 9,249 user–system conversations collected from a conversational search and recommendation platform. Users interact with the system to explore unfamiliar topics and identify products suitable for their needs. Each conversation is supported by a catalogue of 12,312,760 products from the Amazon US platform (May 1996 – September 2023), enriched with item metadata and 303,900,435 user reviews. Additionally, 113,520,750 open-domain textual documents from the Bing search engine (MS-MARCO v2.1) provide general knowledge to contextualize recommendations.

The dataset is further enriched with extensive annotations collected from expert human annotators, including conversation quality assessments, user intent labels, contextualized request annotations, user profile information, and personalized relevance judgements. CoSRec is composed of 9,249 conversations, subdivided in three distinct partitions based on the depth of annotations collected. CoSRec-Raw comprises 8,938 non-annotated conversations, while the remaining 311 are manually annotated and form the CoSRec-Crowd and CoSRec-Curated partitions. These conversations are evaluated among four quality aspects and are labelled to identify user intents in each utterance (4,411 in total). Based on the quality assessment results, 20 high-quality conversations are selected and refined to form the CoSRec-Curated partition, including 17,464 manually crafted personalized relevance judgments.

CoSRec comprises a catalogue of 12,312,760 items associated with 303,900,435 user reviews obtained by filtering the Amazon Reviews 2023 dataset, which has been collected and publicly released by the UC San Diego University under the MIT license. Additionally, CoSRec leverages MS-MARCO v2.1 general open-domain textual data, publicly released by Microsoft for non-commercial research purposes.

From a food safety perspective, CoSRec offers potential applications for mining consumer behaviour, information needs, and product-related discussions, which can help identify emerging risks, patterns in consumer queries about food products, or the prevalence of specific hazards. Its combination of large-scale product metadata, user reviews, and expert-annotated conversations provides a rich resource for developing NLP methods for food safety knowledge extraction, such as identifying frequently mentioned hazards, contamination concerns, or consumer-reported product issues.

#### **Data Processing and Curation**

The dataset aggregates heterogeneous data from multiple sources. Recommendation data is provided in a loosely structured format, with incomplete records and inconsistencies in product metadata across elements. Web-sourced textual documents contain noise, including residual HTML tags, due to automatic extraction processes. To address these challenges, a data preprocessing pipeline is required to enforce a uniform structure, normalize metadata fields, and clean textual content.

Human annotation data are collected from expert annotators and undergo review procedures to mitigate subjectivity and ensure internal consistency across annotations.

#### **Metadata and Content**

The catalogue of items follows a strict custom schema, defined as follows:



- **Parent\_asin:** The unique product identifier.
- **Title:** The name of the product.
- **Description:** The textual description of the product.
- **Details:** The key-value pairs including additional data about the product, such as material, brand, ...
- **Categories:** The hierarchical multi-level categories associated with the product (*i.e.*, Video Games / PC / Accessories / Gaming Keyboards).
- **Store:** The name of the store within Amazon e-commerce which sells the product.
- **Price:** The price in US dollars (\$) of the product.
- **Num\_ratings:** The total number of reviews/ratings given by users to the product, as reported in the product page on Amazon.
- **Avg\_ratings:** The average rating (ranging from 1.0 to 5.0) given by users to the product, as reported in the product page on Amazon.
- **Valid\_reviews:** The list of all reviews available for the product, discarding malformed or empty records.
- **Valid\_ratings:** The number of 1/2/3/4/5 ratings included in the “valid” reviews.
- **Num\_valid\_ratings:** The total number of “valid” reviews/ratings available for the product.
- **Avg\_valid\_ratings:** The average rating (ranging from 1.0 to 5.0) given by users to the product, considering the “valid” reviews/ratings.

Each item is associated with user reviews (303,900,435 in total), which are provided in the “valid\_reviews” product field. The user reviews data adheres to the following custom schema:

- **User:** The unique anonymous identifier of the user which has posted the review.
- **Timestamp:** The timing data (unix time) when the review has been posted by the user.
- **Verified\_purchase:** The boolean (yes/no) value indicating whether the user has purchased the product through Amazon e-commerce.
- **Title:** The title of the user review.
- **Text:** The textual content of the user review.
- **Rating:** The rating score (1/2/3/4/5) given to the product by the user.
- **Helpful:** The number of other users within Amazon e-commerce rating the review as useful and helpful.

The textual documents included in the corpus comprise two fields:

- **Id:** The unique document identifier.
- **Content:** The textual content of the document.

In CoSRec, multiple annotations are gathered from expert human annotators, including conversation quality assessments, user intent labels, contextualized request annotations, user profile information, and personalized relevance judgements. All the annotation data collected for the CoSRec dataset are distributed as a list of conversation id-content pairs.

The conversations are provided as is. Within the conversation text, the start of every user and system turns are signalled by “U: ” and “S: ”, respectively.

The conversation quality assessment data comprises the list of judgments provided by up to 6 distinct human annotators to every conversation, using a 1 to 5 scale, according to 4 quality criteria: fluency, informativeness, logicity, and coherence.

The user intent labelling enriches each user turn to determine their implicit and explicit requests embedded in the utterance. Such data is organized as a list of records following this schema:

- **Utterance:** The (0-based) user turn number in which the following intent(s) is/are extracted from.
- **Intents:** The list of all intents, which may be 0, 1, or more than 1, identified by the human assessors. Each intent in this list is structured as follows:
- **Id:** The unique intent identifier.

- **Type:** The type associated with the intent, discriminating between “search”, “recommendation”, and “product details” intents.
- **Query\_variants:** The list of formulations expressing the information/product request expressed or implied by the user.
- **Product:** The identifier of the product which the intent refers to (used only for “product details” intents).

Each conversation is associated with 3 or 4 different user profiles, describing the main user characteristics and important facts expressed in their prior reviews. Such data guides the document retrieval and product recommendation by the system, along with the personalized relevance judgments. The data is structured as a dictionary of users id, each associated with the corresponding profile text.

The personalized relevance judgment consists in a list of documents/items assessed by human annotators to establish their degree of relevance for the request implied by every intent for the specific user considered. The data is organized as a list of records, following the standard TREC format (“intent\_id 0 product\_id/item\_id degree\_of\_relevance”).

All data included in the dataset is exclusively in the English language. All data is provided in JSONL format, except for the relevance judgements which adhere to the standard TREC format.

### **Availability and Licensing**

The dataset is provided on Github<sup>4</sup>. All conversations and the data for human annotations collected can be found in the repository. Instead, the document corpus and product catalogue can be built using python scripts provided in the Github repository, starting from the MS-MARCO<sup>5</sup> and Amazon<sup>6</sup> Reviews 2023 original data.

### **Contribution to project KPIs**

**Table 38** presents the contribution of the regulations from CosRec dataset to the data-source-related KPIs, showing how this dataset supports EFRA’s data-size monitoring and KPI target assessment.

**Table 38:** Contribution of the CosRec dataset to the data source related KPIs

KPI	Value	Description
1.2 Sources	3	Data sources
1.3 Aggr. Data Vol.	9,249	Data records
1.4 Soc. & News Data	9,249	Social posts
1.11 Langs Supp.	1	Natural languages

## 7.2. The CompreHensive European Food Safety (CHEFS) Database

### **Overview**

The CompreHensive European Food Safety (CHEFS) database is a harmonized compilation of official food and feed monitoring data submitted annually by EU Member and associated states to the European Food Safety Authority (EFSA) under the Standard Sample Description reporting framework. EFRA collaborated with the HOLiFOOD and ECO-Ready projects to deliver this database and relevant reporting. CHEFS aggregates

<sup>4</sup> <https://github.com/CAMEO-22/CoSRec>.

<sup>5</sup> [https://msmarco.z22.web.core.windows.net/msmarcoranking/msmarco\\_v2.1\\_doc\\_segmented.tar](https://msmarco.z22.web.core.windows.net/msmarcoranking/msmarco_v2.1_doc_segmented.tar)

<sup>6</sup> <https://amazon-reviews-2023.github.io/>



analytical results on chemical contaminants, pesticide residues, and veterinary medicinal product residues (VMPR) for over 4,000 food and feed products, covering the period 2000–2024. The database contains 392,269,911 analytical results from 15,176,473 samples reported by 35 countries. Approximately 78.2% of results concern pesticide residues, 20.7% VMPR, and 1.1% chemical contaminants, spanning 3,180 unique contaminants and 4,035 product identifiers. For EFRA, CHEFS provides a high-granularity evidence base to support AI models for hazard prediction, early warning, and policy analysis, combining detailed measurements, sampling strategies, and standardized identifiers in one interoperable resource. The database is owned by WFSR.

### **Data Processing and Curation**

CHEFS is built through an automated processing pipeline that downloads all relevant EFSA Zenodo CSV files, harmonizes their structure, and imports them into a relational PostgreSQL schema. The pipeline resolves heterogeneity across hazard domains and years (e.g. differing column names, missing variables, and format changes within VMPR and other groups), separates sample-level and analytical-result-level variables, links SSD1/SSD2 catalogues, and removes duplicates. Validation compares exports from CHEFS back against the original Zenodo files to ensure that no analytical results or metadata are altered or lost, and the authors provide Python scripts (using `psycpg2`) for repeatable extraction and merging into analysis-ready datasets.

CHEFS relies on EFSA's Standard Sample Description (SSD/SSD2) logical model, which defines data elements, controlled vocabularies, and business rules for food safety occurrence data. Contaminants are coded using the PARAM controlled terminology, which provides hierarchical names that CHEFS parses to derive high-level ontology groups (e.g., heavy metals, mycotoxins, toxins), enabling grouping and visualization at concept-cluster level. Food and feed products are represented using EFSA's FoodEx and FoodEx2 classification systems, and the team add a custom grouping layer that maps detailed product codes into broader categories such as "Fruits," "Cereals," "Animal meat and tissues," and "Milk and dairy products."

The construction of CHEFS also includes an explicit assessment of data sparsity and missingness, with missing value rates calculated for each variable to understand which SSD fields are robust enough for analysis. Over time, metadata quality improves, exemplified by a roughly tenfold reduction in samples with unknown origin country between 2017 and 2024. The validation step, in which exports from CHEFS are cross-checked against the EFSA Zenodo inputs, further supports metadata integrity by confirming that harmonization and schema changes do not introduce loss or distortion of information.

### **Metadata and Content**

CHEFS primarily standardize key descriptors such as contaminants and products through controlled terminologies (PARAM, FoodEx/FoodEx2), which use code-based, ontology-like labels rather than free text, effectively normalizing linguistic variation across reporting countries. Country names and codes follow standardized EFSA catalogues, and the database is oriented toward a common technical language rather than storing multiple natural-language versions of labels. As a result, while the resource is not explicitly multilingual in the sense of parallel language fields, it is intrinsically language-neutral and interoperable across EU member states via shared controlled vocabularies.

The CHEFS database is implemented as a PostgreSQL relational database in which each analytical result is linked to exactly one sample, realizing a one-to-many sample-to-result data model. To deal with a large number of variables and high sparsity, the schema separates "core" tables that store essential identifiers and key attributes for samples and analytical results from "rest" tables that hold the remaining SSD variables, and it offers predefined views to simplify common joins and selections. Data are originally ingested from EFSA CSV files, and Python scripts are provided for extracting subsets into flat files or analysis data frames, preserving referential integrity and SSD semantics.

At minimum, the core tables must include identifiers for sample and analytical result; contaminant/hazard code; product/matrix code; sampling country and origin country; dates for sampling and analysis; result value and unit; LOQ; evaluation code (compliance); hazard group (pesticide, chemical contaminant, VMPR); and basic sampling design fields (e.g., objective/selective/suspect/other), plus a small set of technical method fields. Taken together, this implies on the order of 20–40 core columns across the main sample and result



tables that are systematically populated and sufficient to reproduce the analyses in the paper, while the remaining SSD variables are delegated to “rest” tables with much higher sparsity.

### **Availability and Licensing**

The underlying EFSA monitoring datasets used to construct CHEFS are openly available on Zenodo under EFSA’s data reuse conditions, and the CHEFS database itself is described in an open-access article published under a Creative Commons Attribution (CC BY) licence. The authors provide scripts and database structure openly on GitHub<sup>7</sup> to enable reconstruction or local deployment of CHEFS from the EFSA Zenodo releases. Within the EFRA Data Hub context, CHEFS can therefore be exposed as a derived, open resource that mirrors the PostgreSQL schema, with access governed by EFRA’s policies but underpinned by EFSA’s open data and the paper’s CC BY licensing, allowing broad reuse for research and innovation.

### **Contribution to project KPIs**

**Table 39** presents the contribution of the CHEFS database to the data-source-related KPIs, showing how this database supports EFRA’s data-size monitoring and KPI target assessment.

**Table 39:** Contribution of the CHEFS database to the data source related KPIs

KPI	Value	Description
1.2 Sources	1	Data sources
1.3 Aggr. Data Vol.	392,269,911	Data records
1.8 Reg. & Legis. Data	392,269,911	Analytical results
1.9 Sust. & Animal Health Data	392,269,911	Analytical results

<sup>7</sup> <https://github.com/WFSRDataScience/CHEFS>

## 8. Data Assets

This chapter summarizes the findings related to the KPIs associated with data assets collected, generated, curated, or integrated throughout the EFRA project. Table 40 presents all data sources described in the previous sections, highlighting the growth in their size across the project milestones (M15, M24, and M36). While

Table 41 presents the aggregated KPI values per data, including the total achieved values, the defined targets, and the resulting difference.

Overall, the results demonstrate that EFRA has successfully assembled a large-scale, diverse, and multi-modal data ecosystem that substantially exceeds the majority of the predefined KPI targets. The total aggregated data volume reached over 719 million records, surpassing the overall target for aggregated data volume (KPI 1.3) by a significant margin. This confirms the project's capacity to acquire and manage data at scale across heterogeneous sources relevant to food safety.

With respect to scientific and expert knowledge (KPI 1.5), the achieved values notably exceed the target, reflecting the successful integration of curated datasets, systematic review abstracts, expert-annotated corpora, and validation datasets. These assets provide a strong evidence base to support advanced analytics, knowledge extraction, and AI-driven research within the food safety domain.

Similarly, environmental and climatic data (KPI 1.6), economic and trade data (KPI 1.7), regulatory and legislative data (KPI 1.8), and sustainability and animal health data (KPI 1.9) all significantly exceed their respective targets. This outcome highlights the comprehensive coverage achieved across critical dimensions of food safety risk assessment, spanning upstream production factors, trade and market dynamics, regulatory frameworks, and animal health considerations.

The only KPI not fully achieved is KPI 1.4 (Social and News Data). This shortfall is primarily attributed to external constraints related to the availability of social media data, notably changes in access policies for Twitter/X data that occurred after the KPI targets were defined. Despite this limitation, the project still integrated substantial volumes of news and media content from alternative sources, partially mitigating the impact on overall coverage.

In addition, the temporal evolution of data volumes across milestones (M15, M24, and M36) demonstrates continuous growth and enrichment of the EFRA data assets over time. This progression reflects sustained data acquisition efforts, onboarding of new providers, expansion of existing datasets, and the inclusion of newly generated and annotated resources during the later phases of the project.

In conclusion, the KPI analysis confirms that EFRA has met its data-related objectives to a very high degree. The project has delivered a robust and well-balanced portfolio of data assets that not only fulfils but, in most cases, significantly exceeds the defined targets, thereby establishing a strong foundation for food safety research, analytics, and future exploitation beyond the project lifetime.

**Table 40:** EFRA Data Assets (name, data owner, number of records)

Data Source (Section)	Data Owner	Records (by M15)	Records (by M24)	Records (by M36)
Food Product Recalls (6.1.1)	Agroknow	94,994	113,992	126,907
Global Food Product Border Rejections (0)	Agroknow	404,095	497,037	598,738
Lab Tests (0)	Agroknow	128,906,112	144,337,845	226,792,288
Inspections (0)	Agroknow	322,886	358,403	501,949
Global Fraud Reports (6.1.15)	Agroknow	17,204	18,236	147,363

Food Safety News (0)	Agroknow	73,527	73,527	95,076
Foodborne Disease Outbreaks (0)	Agroknow	74,134	80,806	4,064 (Provider has reduced the size of the available data)
Weather Data (0)	Agroknow	1,779,418	1,853,441	1,925,339
Poultry Production Data (0)	Agroknow	9,766	9,766	50.090
Salmonella Testing (0 first dataset)	Agroknow	9,766	9,766	9,766
Mycotoxins Testing (0, 0)	Agroknow	6,101	6,101	7,354
Prediction Trends (0)	Agroknow	0	10.318.338	42,992,662
Prediction Risks (0)	Agroknow	0	335.769	3,052,452
Production Data (0)	Agroknow	503,329	503,329	503,329
Trade Data (0)	Agroknow	41,424,465	41,424,465	41,424,465
Price data (0)	Agroknow	0	0	415,670
Food Recall Incidents (4.1)	Agroknow	7,546	7,546	7,546
Weather Data (6.2.1)	Agrivi	1,085	1,917	2,000,000
Pest Observations (0)	Agrivi	342	887	1,613
SGS Digicomply Dataset (4.1)	SGS Digicomply	14,308	14,308	127,204
Chunk Relevance Validation Dataset (4.2)	SGS Digicomply	0	0	1,330
FoodSafeSum Dataset (4.3)	SGS Digicomply	0	0	2,091
Abstracts of a systematic review on Food and federated learning (0)	WFSR	0	0	548
Abstracts of a systematic review on Food and NLP (0)	WFSR	0	0	244
News Articles by WFSR (0)	WFSR	0	0	2,000
CHEFS (0)	WFSR	0	0	392,269,911
CoSRec (7.1)	CNR	0	0	9,249

**Table 41:** Aggregated KPI values per data source, including totals, targets, and differences

KPI/Owner	Data Source	1.3 AGGR. DATA VOL.	1.4 SOC. & NEWS DATA	1.5 SCI. & EXPERT KNOWL.	1.6 ENV. & CLIM. DATA.	1.7 ECON. & TRADE DATA	1.8 REG. & LEGIS. DATA	1.9 SUST. & ANIMAL HEALTH DATA
Agrivi	Weather Data	2,000,000	0	0	2,000,000	0	0	0
	Pest Observations	1,613	0	0	1,613	0	0	1,613

Agrokno w	Food Product Recalls	126,907	0	0	0	0	126,907	0
	Global Food Product Border Rejections	598,738	0	0	0	598,738	598,738	0
	Lab Tests	226,792,288	0	0	226,792,288	0	0	226,792,288
	Inspections	501,949	0	0	0	0	501,949	501,949
	Global Fraud Reports	147.363	1,887	0	0	0	145,476	0
	Food Safety News	95,076	95,076	0	0	0	0	0
	Foodborne Disease Outbreaks	4,064	0	0	0	0	4,064	0
	Weather Data	1,925,339	0	0	1,925,339	0	0	0
	Mycotoxins Testing (Food Fortress Dataset)	864	0	864	0	0	0	0
	Prediction Trends	42,992,662	0	0	42,992,662	42,992,662	42,992,662	0
	Prediction Risks	3,052,452	0	0	3,052,452	0	3,052,452	3,052,452
	Production Data	503,329	0	0	0	503,329	0	0
	Trade Data	41,424,465	0	0	0	41,424,465	0	0
	Price Data	415,670	0	0	0	415,670	0	0
	Poultry Production Data	50,090	0	0	0	0	0	50,090
	Food Recall Incidents	7,546	0	0	0	0	7,546	0
CNR	CoSRec	9,249	9,249	0	0	0	0	0
SGS Digicomply	SGS Digicomply Dataset	127,204	80,504	32,672	0	0	46,454	0
	Chunk Relevance Validation Dataset	1330	96	24	0	0	25	0

	FoodSafeSum Dataset	2,091	1,320	419	0	0	640	0
WFSR	Abstracts of a systematic review on Food and federated learning	548	0	548	0	0	0	0
	Abstracts of a systematic review on Food and NLP	244	0	244	0	0	0	0
	News Articles by WFSR	2,000	2,000	0	0	0	0	0
	CHEFS	392,269,911	0	0	0	0	392,269,911	392,269,911
	Total	712,402,087	190,132	33,907	276,764,354	85,934,864	439,746,824	622,668,303
	Target	600,000,000	200,000,000	15,000	2,000,000	3,000,000	2,000,000	1,000,000
	Difference (Target – Total)	--112,402,087	199,809,868	-18,907	--274,764,354	--82,934,864	---437,746,824	--621,668,303

## 9. Conclusions

This deliverable has presented a comprehensive and structured overview of the data assets underpinning the EFRA project, covering primary, external, industrial, and composite datasets relevant to food safety, agriculture, and decision-support applications. Through systematic documentation, the deliverable establishes a clear picture of the scope, provenance, structure, and accessibility of the data sources that feed EFRA's analytical and AI-driven components.

The work demonstrates that EFRA has access to a diverse and complementary portfolio of datasets, ranging from large-scale, high-granularity monitoring data (e.g., CHEFS, weather and environmental data) to document-centric resources (e.g., legislation, scientific literature, news and reports) and complex composite datasets that integrate multiple data modalities. Despite significant heterogeneity in scale, structure, and licensing conditions, these data sources can be effectively integrated through the use of standardized schemas, controlled vocabularies, well-defined processing pipelines, and consistent documentation practices.

A key conclusion is that the majority of the documented datasets are analytically ready or can be made so with limited additional processing, enabling their direct use in EFRA pilots and downstream AI workflows. The explicit identification of data volumes, update frequencies, metadata fields, and processing rules supports reproducibility, transparency, and reliable KPI tracking across the project. At the same time, the deliverable highlights practical constraints, such as restricted licensing or commercial access for certain datasets, which must be managed at the pilot and exploitation stages.

The analysis of data related KPIs further confirms that EFRA has met or exceeded its targets across nearly all categories, validating the adequacy, scale, and balance of the data assets assembled throughout the project. Overall, this deliverable establishes a shared and authoritative reference on EFRA's data foundations, supporting alignment among partners and providing reviewers and stakeholders with a clear understanding of how data resources underpin the project's objectives. By consolidating knowledge on data availability, quality, and integration pathways, the deliverable lays a solid foundation for the implementation, evaluation, and long-term sustainability of EFRA's food safety and decision-support services.

## Annex I

This annex provides a detailed overview of the data sources and datasets collected in the project, supporting the traceability of KPI 1.1, KPI 1.2 and KPI 1.11 values. Table 42 summarizes the data types considered, the corresponding data sources, and their owners. Each row corresponds to a unique data type ID, allowing clear identification and reference in subsequent analyses. For KPI 1.1, the target value was 20, and the project collected 21 diverse data types, exceeding the target.

Table 43 lists the sources of the datasets provided by the various data owners that contribute to KPI 1.2. For each source, the table identifies the relevant datasets and their corresponding owners. This mapping provides a clear view of which sources support KPI 1.2 and facilitates traceability from the datasets to the KPI values. A total of 686 diverse data sources were collected, exceeding the KPI 1.2 target of 400.

Table 44 presents the languages supported by the datasets provided by the various data owners, contributing to KPI 1.11. For each language, the table lists the language code and name, the associated datasets, and the respective data owners. This overview allows for an understanding of the linguistic coverage of the datasets and their support for KPI 1.11. In total, 62 diverse languages were collected, surpassing the KPI 1.11 target of 10.

**Table 42:** The data types of the data sources provided by the owners that support KPI 1.1.

ID	Data Type	Data Sources	Data Owners
1	Food Product Recalls	Food Product Recalls, Product Recall Incidents	Agroknow
2	Food Product Border Rejections	Food Product Border Rejections	Agroknow
3	Lab Tests	Lab Tests	Agroknow
4	Inspections	Inspections	Agroknow
5	Fraud Reports	Global Fraud Reports	Agroknow
6	Food Safety News	Food Safety News, SGS Digicomply Dataset, Chunk Relevance Validation Dataset, FoodSafeSum, News Articles by WFSR	Agroknow, SGS Digicomply, WFSR
7	Foodborne Disease Outbreaks	Foodborne Disease Outbreaks	Agroknow
8	Weather Data	Weather Data	Agroknow, Agrivi
9	Poultry Farm Production	Poultry Farm Production	Agroknow
10	Salmonella Testing	Salmonella Testing	Agroknow
11	Mycotoxins Testing	Mycotoxins Testing	Agroknow
12	Prediction Trends	Prediction Trends	Agroknow
13	Prediction Risks	Prediction Risks	Agroknow
14	Production Data	Production Data	Agroknow
15	Trade Data	Trade Data	Agroknow
16	Pricing Data	Pricing Data	Agroknow
17	Pest Observations	Pest Observations	Agrivi
18	Scientific Publications	SGS Digicomply Dataset, Chunk Relevance Validation Dataset, FoodSafeSum, Abstracts of a systematic review on Food and	SGS Digicomply

		federated learning, Abstracts of a systematic review on Food and NLP	
19	Regulations	SGS Digicomply Dataset, Chunk Relevance Validation Dataset, FoodSafeSum	SGS Digicomply
20	Guidance	SGS Digicomply Dataset, Chunk Relevance Validation Dataset, FoodSafeSum	SGS Digicomply
21	Social media	CosRec	CNR
22	Food and Feed Monitoring Data	CHEFS	WFSR

**Table 43:** The sources of the data sources provided by the owners that support KPI 1.2

ID	Source Name	Datasets	Owners
1	ACL Anthology	Abstracts of a systematic review on Food and NLP	WFSR
2	AGES (Austria)	Food Recall Incidents, Food Recalls	Agroknow
3	Australian Product Safety	Food Recall Incidents	Agroknow
4	BBC	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
5	CAB Abstracts	Abstracts of a systematic review on Food and federated learning, Abstracts of a systematic review on Food and NLP	WFSR
6	CNN	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
7	Canadian Food Inspection Agency (CFIA)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
8	Centre for Food Safety (Hong Kong)	Food Recall Incidents, FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
9	Danish Veterinary and Food Administration	Food Recall Incidents, Food Recalls	Agroknow
10	EUR-Lex	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
11	European Food Safety Authority (EFSA)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Lab Tests, CHEFS	SGS Digicomply, Agroknow, WFSR
12	European Food Safety Authority (EFSA) Publications	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
13	Food Standards Australia New Zealand (FSANZ)	Food Recall Incidents, FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
14	Food Standards Scotland	Food Recall Incidents, FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
15	France 24	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
16	Google Scholar	Abstracts of a systematic review on Food and federated learning	WFSR



17	Government of Canada	Food Recall Incidents, FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	Agroknow, SGS Digicomply
18	IEEE Xplore	Abstracts of a systematic review on Food and federated learning, Abstracts of a systematic review on Food and NLP	WFSR
19	Inspection Canada	Food Recall Incidents, FoodSafeSum, SGS Digicomply Dataset	Agroknow, SGS Digicomply
20	Lebensmittelwarnung.de	Food Recall Incidents, Food Recalls	Agroknow
21	Publications Office of the European Union	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
22	Reuters	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
23	Scopus	Abstracts of a systematic review on Food and federated learning, Abstracts of a systematic review on Food and NLP	WFSR
24	Singapore Food Agency	Food Recall Incidents, Food Recalls	Agroknow
25	The Guardian	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
26	The New York Times	FoodSafeSum, SGS Digicomply Dataset, News Articles by WFSR	SGS Digicomply, WFSR
27	U.S. Food and Drug Administration (FDA)	Food Recall Incidents, FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Food Recalls, Foodborne Disease Outbreaks, Inspections	SGS Digicomply, Agroknow
28	UK Food Standards Agency (FSA)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
29	UK Government	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
30	USDA Food Safety and Inspection Service (FSIS)	Food Recall Incidents, FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	Agroknow, SGS Digicomply
31	Web of Science	Abstracts of a systematic review on Food and federated learning, Abstracts of a systematic review on Food and NLP	WFSR
32	acc product safety	Border Rejections	Agroknow
33	acs environmental science and technology	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
34	act government	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
35	action on salt	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
36	actiononsugar.org	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
37	actu-environnement.com	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
38	administración nacional de medicamentos, alimentos y tecnología médica	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply

39	Administration luxembourgeoise vétérinaire et alimentaire	Food Recalls	Agroknow
40	advertising standards authority ltd. (asa)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
41	afepadi	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
42	afn (agfundernews)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
43	ag week	SGS Digicomply Dataset	SGS Digicomply
44	agence nationale de sécurité du médicament et des produits de santé	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
45	agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (anses)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
46	agencia estatal boletín oficial del estado - spain	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
47	agency for standardization, metrology, certification and trade inspection of the republic of tajikistan "tajikstandard"	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
48	agri-food and biosciences institute (afbi)	SGS Digicomply Dataset	SGS Digicomply
49	agriculture and agri-food canada - news - government of canada	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
50	agrifoodtoday	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
51	agriland	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
52	agro-messages abroad	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
53	agrobond	SGS Digicomply Dataset	SGS Digicomply
54	agropages.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
55	alim'agri - site du ministère de l'agriculture et de l'alimentation	SGS Digicomply Dataset	SGS Digicomply
56	all about feed	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
57	allergen bureau food	SGS Digicomply Dataset	SGS Digicomply
58	alnilin.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
59	alta soft	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
60	amazon reviews 2023 dataset	CoSRec	CNR
61	amazon us product catalogue	CoSRec	CNR
62	american college of allergy, asthma & immunology	SGS Digicomply Dataset	SGS Digicomply
63	american pet products association (appa)	SGS Digicomply Dataset	SGS Digicomply
64	andean community	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

65	andean community - can	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
66	ania	SGS Digicomply Dataset	SGS Digicomply
67	animal feed and veterinary products control - thailand	SGS Digicomply Dataset	SGS Digicomply
68	aoac international	SGS Digicomply Dataset	SGS Digicomply
69	argentina.gob.ar presidencia de la nación	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
70	arla foods ingredients	SGS Digicomply Dataset	SGS Digicomply
71	asociación de consultores y formadores de españa en calidad y seguridad alimentaria	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
72	asociación nacional de fabricantes de productos de dietética infantil (andi)	SGS Digicomply Dataset	SGS Digicomply
73	association of dutch manufacturers of children's and dietetic foods (vnfkd)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
74	association of national organizations of fishing enterprises in the european union	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
75	association of southeast asian nations (asean)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
76	association of the european adhesive and sealant industry (feica)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
77	assoenologi (the association of oenologists and italian winemakers)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
78	australian competition and consumer commission	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
79	australian food news	SGS Digicomply Dataset	SGS Digicomply
80	australian government - australian trade and investment commission	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
81	australian government - federal register of legislation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
82	australian government department of agriculture and water resources	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
83	australian government department of health	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
84	australian packaging covenant organisation (apco)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
85	australian pesticides and veterinary medicines authority	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
86	azti science and technology centre	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

87	baby milk action	SGS Digicomply Dataset	SGS Digicomply
88	bangladesh food safety authority	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
89	bayer global	SGS Digicomply Dataset	SGS Digicomply
90	bbb national programs newsroom, national advertising division	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
91	belarus news (belta)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
92	belta.by	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
93	beyond pesticides daily news blog	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
94	bioforum	SGS Digicomply Dataset	SGS Digicomply
95	bloomberg law	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
96	boletin oficial republica argentina	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
97	borneo bulletin online	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
98	bpjph (badan penyelenggara jaminan produk halal)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
99	brazil official diary of the union	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
100	brazilian health regulatory agency - anvisa	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
101	british crop production council (bcpc)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
102	british journal of sports medicine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
103	bulgaria – държавен вестник - state gazette - regulation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
104	bulgarian food safety agency	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
105	bulletin officiel du ministère de l'agriculture et de l'alimentation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
106	bundesamt für verbraucherchutz und lebensmittelsicherheit	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
107	bundesinstitut für risikobewertung	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
108	bureau of food - food and drug administration (fda) thailand	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
109	bureau of standards jamaica	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
110	california department of justice	SGS Digicomply Dataset	SGS Digicomply
111	california department of public health	Food Recalls	Agroknow
112	california legislative information	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
113	california office of environmental health hazard assessment (oehha)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
114	california rice news	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

115	cambodia general department of consumer protection, competition and fraud suppression	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
116	cdc	Foodborne Disease Outbreaks	Agroknow
117	center for food safety	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
118	centers for disease control and prevention (cdc)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
119	cgnt	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
120	chartered institute of environmental health (cieh)	SGS Digicomply Dataset	SGS Digicomply
121	chem trust	SGS Digicomply Dataset	SGS Digicomply
122	chemical watch	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
123	chemlinked	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
124	chemsec	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
125	chestny znak state track & trace system	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
126	chilled food association (cfa)	SGS Digicomply Dataset	SGS Digicomply
127	china briefing	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
128	china national center for food safety risk assessment	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
129	china national food quality supervision and inspection center	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
130	china national health commission	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
131	china national standardization centre of food & fermentation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
132	china quality news network	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
133	cirad	SGS Digicomply Dataset	SGS Digicomply
134	clcv (national association for the defense of consumers and users)	SGS Digicomply Dataset	SGS Digicomply
135	climate.nasa.gov	SGS Digicomply Dataset	SGS Digicomply
136	cms law now	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
137	cna - channel news asia	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
138	coceral - european code of good trading practice	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
139	codex alimentarius international food standards	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
140	cokz.nl	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
141	comisión federal para la protección contra riesgos sanitarios - federal commission for protection against health risks	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

142	confectionary industry news	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
143	consilium europa	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
144	consultations.dh.gov.uk	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
145	consumer affairs agency, government of japan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
146	consumer federation of america	SGS Digicomply Dataset	SGS Digicomply
147	consumer reports	SGS Digicomply Dataset	SGS Digicomply
148	consumer reports, inc	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
149	corteva agriscience	SGS Digicomply Dataset	SGS Digicomply
150	court of justice of the european union	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
151	courthouse news service	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
152	crop and pasture science	SGS Digicomply Dataset	SGS Digicomply
153	croplife	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
154	csa group	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
155	csp daily news	SGS Digicomply Dataset	SGS Digicomply
156	cyprus mail	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
157	database of gosts	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
158	defend our health	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
159	denmark retsinformation	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
160	denmarks tekniske universitet - technical university of denmark	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
161	department for environment, food and rural affairs	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
162	department of agriculture - thailand	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
163	department of agriculture, land reform and rural development (dalrrd)	SGS Digicomply Dataset	SGS Digicomply
164	deputy camara - chile	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
165	diario oficial de la federación (dof)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
166	dietary guidelines for americans (dga)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
167	direção-geral de alimentação e veterinária	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
168	dnt (data network for technologies)	Pest Observations	Agrivi
169	documentation and legal information networks of the ministry of industry	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
170	downey brand	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

171	drinkstuff sa	SGS Digicomply Dataset	SGS Digicomply
172	dutchnews.nl	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
173	e-contrast blogspot	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
174	e-gov japan	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
175	e-people south korea	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
176	ecetoc	SGS Digicomply Dataset	SGS Digicomply
177	edizioni pubblicità italia	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
178	efa news - european food agency	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
179	egyptian national food safety authority	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
180	el país	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
181	el peruano	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
182	electronic code of federal regulations - u.s.	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
183	electronic fund legal and regulatory technical documents	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
184	electronic irish statute book (eish)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
185	en.unesco.org	SGS Digicomply Dataset	SGS Digicomply
186	energy & climate intelligence	SGS Digicomply Dataset	SGS Digicomply
187	environmental defense fund	SGS Digicomply Dataset	SGS Digicomply
188	environmental services association (esa)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
189	epthinktank.eu	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
190	ervin cohen & jessup llp	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
191	esg today	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
192	esm - the european supermarket magazine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
193	estonian agricultural board	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
194	eu reporter	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
195	euractive	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
196	eurasian economic commission (eec)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
197	eurasian economic union legal portal	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
198	eurekaalert	SGS Digicomply Dataset	SGS Digicomply
199	euromilk	SGS Digicomply Dataset	SGS Digicomply
200	european association of fruit and vegetable processors	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply



201	european centre for disease prevention and control	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
202	european chemicals agency	SGS Digicomply Dataset	SGS Digicomply
203	european commission	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
204	european commission comitology register	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
205	european commission press release database	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
206	european commission scientific committees on consumer safety (sccs)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
207	european environment agency	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
208	european environment bureau	SGS Digicomply Dataset	SGS Digicomply
209	european free trade association (efta)	SGS Digicomply Dataset	SGS Digicomply
210	european parliament	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
211	european public health alliance	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
212	european union agriculture and rural development agricultural	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
213	european union joint research centre (jrc)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
214	european union observatory for nanomaterials (euon)	SGS Digicomply Dataset	SGS Digicomply
215	europol	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
216	evira finnish food safety authority - product recalls	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
217	faolex database	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
218	fare	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
219	farminguk	SGS Digicomply Dataset	SGS Digicomply
220	fda philippines	Food Recalls	Agroknow
221	federal agency for the safety of the food chain (fasfc)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
222	federal customs service of russia	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
223	federal food safety and veterinary office	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply



224	federal portal of draft regulatory legal acts	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
225	federal service for supervision in the sphere of protection of consumers 'rights and human rights	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
226	federal trade commission (ftc)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
227	federation of german food and drink industries (bve)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
228	federation of the dutch food industry (fnli)	SGS Digicomply Dataset	SGS Digicomply
229	fevia	SGS Digicomply Dataset	SGS Digicomply
230	fi global insights	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
231	finnish food authority	Food Recalls	Agroknow
232	fipa	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
233	focus taiwan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
234	food & beverage industry news (australia)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
235	food & beverage news	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
236	food & drink federation - news	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
237	food allergy canada	SGS Digicomply Dataset	SGS Digicomply
238	food allergy research & education (fare)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
239	food and agriculture organization of the united nations (fao)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, Production Data, Trade Data, Pricing Data	SGS Digicomply, Agroknow
240	food and drink manufacture	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
241	food and drug administration philippines	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
242	food authenticity network	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
243	food business news	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
244	food dive	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
245	food fraud prevention think tank	SGS Digicomply Dataset	SGS Digicomply
246	food in canada	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
247	food ingredients first	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

248	food ingredients online	SGS Digicomply Dataset	SGS Digicomply
249	food law consult	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
250	food magazine   south west food and drink	SGS Digicomply Dataset	SGS Digicomply
251	food manufacturing - advantage business media	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
252	food navigator-asia.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
253	food navigator.com	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
254	food online	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
255	food packaging forum	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
256	food partner network - 食品伙伴网	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
257	food poison journal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
258	food processing australia	SGS Digicomply Dataset	SGS Digicomply
259	food regulation secretariat	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
260	food safety africa	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
261	food safety agency of the republic of azerbaijan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
262	food safety and quality division   ministry of health malaysia	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
263	food safety and standards authority of india	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
264	food safety and standards authority of india food import rejection alert (fira)	Border Rejections	Agroknow
265	food safety authority of hungary	Food Recalls	Agroknow
266	food safety authority of ireland	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
267	food safety authority of luxembourg	Food Recalls	Agroknow
268	food safety authority of vietnam	Food Recalls	Agroknow
269	food safety commission of japan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
270	food safety institute vietnam	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
271	food safety magazine	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply

272	food safety news	FoodSafeSum, SGS Digicomply Dataset, Food Safety News	SGS Digicomply, Agroknow
273	food safety tech	Food Safety News	Agroknow
274	foodconnection - brazil	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
275	fooddrinkeurope	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
276	foodnavigator-usa.com/william reed business media sas	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
277	foodpro	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
278	foodprocessing-technology.com	SGS Digicomply Dataset	SGS Digicomply
279	foodsafety tech	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
280	foodsmi	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
281	foodwatch	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
282	forbes	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
283	francetvinfo.fr	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
284	frankfurt kurnit klein & selz pc (fkks) advertising law firm	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
285	french food observatory	SGS Digicomply Dataset	SGS Digicomply
286	fresh plaza	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
287	friends of the earth	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
288	frontiers in nutrition	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
289	fsa import refusals	Border Rejections	Agroknow
290	future farming	SGS Digicomply Dataset	SGS Digicomply
291	fédération des exportateurs de vins & spiritueux de france	SGS Digicomply Dataset	SGS Digicomply
292	gamberorosso	SGS Digicomply Dataset	SGS Digicomply
293	ganado advocates	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
294	gazzetta ufficiale della repubblica italiana	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
295	general mills: a u.s.-based food company	SGS Digicomply Dataset	SGS Digicomply
296	genetic literacy project	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
297	georgia department of agriculture and commerce	Food Recalls	Agroknow
298	german farmers' association - dbv	SGS Digicomply Dataset	SGS Digicomply
299	german federal ministry of food, agriculture and consumer protection	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply

300	german official gazette publisher	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
301	gezondheid.be	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
302	global panel on agriculture and food systems for nutrition	SGS Digicomply Dataset	SGS Digicomply
303	global witness	SGS Digicomply Dataset	SGS Digicomply
304	globalcompliancenews	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
305	gmp+	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
306	gobierno de la federación	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
307	gouvernement.lu	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
308	gov.si	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
309	government of ireland	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
310	government of quebec	Food Recalls	Agroknow
311	government portal of ukraine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
312	government.nl	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
313	governo do brasil	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
314	govhk portal - hong kong special administrative region government	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
315	grain and feed trade association (gafta)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
316	great italian food trade	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
317	green queen media	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
318	gulf news	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
319	halal malaysian portal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
320	haut conseil des biotechnologies (hcb)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
321	health and environment alliance (heal)	SGS Digicomply Dataset	SGS Digicomply
322	health and safety executive (hse) uk	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
323	health promotion board	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
324	health science authority singapore	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
325	health, food chain safety and environment	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
326	healthcare without harm (hcwh) europe	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
327	healthy caribbean coalition	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
328	hindustan times	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

329	hogar sin toxicos	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
330	hong kong e-legislation (hkel)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
331	how stuff works	SGS Digicomply Dataset	SGS Digicomply
332	i alimentar	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
333	icis	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
334	icta islamabad capital territory administration	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
335	idec - instituto brasileiro de defesa do consumidor	SGS Digicomply Dataset	SGS Digicomply
336	ifoam organics europe	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
337	ifpri : international food policy research institute	SGS Digicomply Dataset	SGS Digicomply
338	ifst	SGS Digicomply Dataset	SGS Digicomply
339	il fatto alimentare	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
340	il sole 24 ore	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
341	ilsi.eu	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
342	india environment portal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
343	india legal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
344	indonesia ministry of religion	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
345	indonesian directorate of processed food standardization	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
346	indonesian fda - badan pengawas obat dan makanan - republik indonesia	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
347	indonesian food and drug monitoring agency	Food Recalls	Agroknow
348	indonesian ministry of trade	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
349	industry and commerce ministry - national information and notification system snin -paraguay	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
350	infobae	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
351	información legislativa y documental - ministerio de justicia y derechos humanos de la nación	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
352	informador.mx	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
353	informační centrum bezpečnosti potravin	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
354	inside climate news	SGS Digicomply Dataset	SGS Digicomply
355	institut national de l'origine et de la qualité (inao)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

356	instituto nacional de metrologia, qualidade e tecnologia	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
357	instituto nacional de salud - national institute of health	SGS Digicomply Dataset	SGS Digicomply
358	insurance journal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
359	international dairy federation	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
360	international nut and dried fruit council	SGS Digicomply Dataset	SGS Digicomply
361	international renewable energy agency	SGS Digicomply Dataset	SGS Digicomply
362	international service for the acquisition of agri-biotech applications (isaaa)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
363	internet legal system database (isap) - parliament of the republic of poland	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
364	investigation offices of baden-wuerttemberg	SGS Digicomply Dataset	SGS Digicomply
365	iom3 (institute of materials, minerals & mining)	SGS Digicomply Dataset	SGS Digicomply
366	irish farmers journal	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
367	iso - international organization for standardization	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
368	israel chamber of commerce	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
369	israel government	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
370	italian ministry of health	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
371	italianfood.net	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
372	jamanetwork.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
373	japanese industrial standards	SGS Digicomply Dataset	SGS Digicomply
374	jd supra	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
375	jordbruks verket - swedish board of agriculture	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
376	journals.plos.org	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
377	keller and heckman	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
378	kommunikationsplattform verbraucher-gesundheit kvg	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
379	korea joongang daily	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

380	korea ministry of food and drug safety (mfds)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
381	korea ministry of government legislation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
382	korean ministry of environment	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
383	korean national fisheries products quality assurance service	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
384	korean national law information center	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
385	krishi jagran	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
386	la tribune.fr	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
387	laws & regulation database of the republic of china (taiwan)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
388	le monde	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
389	le temps	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
390	legal and information system of republic serbia	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
391	legifrance	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
392	legislation of ukraine	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
393	let's do it foundation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
394	lex.uz online	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
395	ley chile	SGS Digicomply Dataset	SGS Digicomply
396	lineaires	SGS Digicomply Dataset	SGS Digicomply
397	livemint	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
398	livsmedelsföretagen (swedish food producers' association)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
399	lovdata - norway legal information system	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
400	lppom mui	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
401	malaysian palm oil council (mpoc)	SGS Digicomply Dataset	SGS Digicomply
402	malta today	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
403	manila bulletin	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
404	manitoba co-operator	SGS Digicomply Dataset	SGS Digicomply
405	mars	SGS Digicomply Dataset	SGS Digicomply



406	matportalen - norway	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
407	mauritius ministry of health and quality of life	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
408	mdpi journals	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
409	media statements	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
410	medisys	Food Safety News	Agroknow
411	michigan state university, institute for food laws and regulations	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
412	mlenio	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
413	miljø-og fødevarestyrelsen - veterinary and food administration	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
414	ministerio de agricultura y pesca, alimentación y medio ambiente	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
415	ministerio de salud y protección social - minsalud	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
416	ministerio della salute	Food Recalls	Agroknow
417	ministero delle politiche agricole alimentari e forestali	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
418	ministero dello sviluppo economico	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
419	ministry for development of economy, trade and agriculture of ukraine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
420	ministry for the environment new zealand	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
421	ministry of agrarian policy and food of ukraine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
422	ministry of agriculture and environmental protection - veterinary inspection	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
423	ministry of agriculture and rural development - israel	SGS Digicomply Dataset	SGS Digicomply
424	ministry of agriculture and rurale development - vietnam	SGS Digicomply Dataset	SGS Digicomply
425	ministry of agriculture of chile	Food Recalls	Agroknow
426	ministry of agriculture of the czech republic	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
427	ministry of agriculture of the people's republic of china	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
428	ministry of agriculture of the russian federation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
429	ministry of agriculture, forestry and fisheries	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply



430	ministry of agriculture, nature and food quality of the netherlands	Food Recalls	Agroknow
431	ministry of climate, change and environment	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
432	ministry of commerce and industry, government of india	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
433	ministry of development, industry and commerce - nicaragua	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
434	ministry of ecology and natural resources of ukraine - regulation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
435	ministry of health of argentina	Food Recalls	Agroknow
436	ministry of health of ukraine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
437	ministry of health singapore	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
438	ministry of health, labour and welfare - japan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
439	ministry of health, labour and welfare japan	Border Rejections	Agroknow
440	ministry of industry and advanced technology of united arab emirates	Food Recalls	Agroknow
441	ministry of justice - government of jamaica	SGS Digicomply Dataset	SGS Digicomply
442	ministry of justice and law	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
443	ministry of justice of the kyrgyz republic	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
444	ministry of public health of chile	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
445	ministry of rural affairs - estonia	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
446	ministry of the environment - japan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
447	ministère de la transition écologique	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
448	ministério da agricultura, pecuária e abastecimento	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
449	modern farmer	SGS Digicomply Dataset	SGS Digicomply
450	mongabay	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
451	ms-marco v2.1 corpus	CoSRec	CNR
452	my ollie	SGS Digicomply Dataset	SGS Digicomply
453	mybusinessmedia bv	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
454	national agency for health regulation, control and surveillance of ecuador	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
455	national assembly of zambia	SGS Digicomply Dataset	SGS Digicomply
456	national center for biotechnology information (ncbi, including pubmed)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

457	national confectioners association (nca)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
458	national herald	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
459	national institute for food and drug surveillance of colombian import refusals	Border Rejections	Agroknow
460	national institute for public health and the environment (rivm)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
461	national institute of food and nutrition - inan - paraguay	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
462	national institute of standards and technology	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
463	national pharmaceutical regulatory agency (npra)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
464	national pig association (npa)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
465	national portal of india	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
466	national post	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
467	national university of singapore	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
468	nature	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
469	nemzeti jogszabálytár kereső - national regulatory finder	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
470	neogen	SGS Digicomply Dataset	SGS Digicomply
471	netherlands food and consumer product safety authority (nvwa)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
472	network documentation and legal information ministry of agriculture of the republic of indonesia	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
473	new food magazine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
474	new hampshire department of health and human services (dhhs)	Food Recalls	Agroknow
475	new hope network	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
476	new south wales government - nsw legislation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
477	new york department of agriculture and markets	Food Recalls	Agroknow
478	new zealand environmental protection authority	SGS Digicomply Dataset	SGS Digicomply
479	new zealand legislation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
480	new zealand ministry of primary industries (mpi)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
481	news-medical.net	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
482	news24	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

483	newsweek	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
484	newswise	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
485	nordic cooperation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
486	norwegian food safety authority mattilsynet	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
487	norwegian institute of public health (niph) - news	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
488	nos	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
489	nrc.nl	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
490	nu.nl	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
491	nutra ingredients	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
492	nutra ingredients usa.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
493	nzherald.co.nz	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
494	office of the united states trade representative	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
495	official gazette - chile	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
496	official internet portal legal information russian federation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
497	olive oil times	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
498	organic consumers association	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
499	organisation for economic co-operation and development	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
500	organismo salvadoreño de reglamentacion tecnica - osartec	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
501	packaging digest	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
502	packaging gateway	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
503	packaging law.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
504	packaging news australia	SGS Digicomply Dataset	SGS Digicomply
505	packaging today	SGS Digicomply Dataset	SGS Digicomply
506	paperity.org	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
507	people's republic of china government portal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
508	pesticide action network europe (pan europe)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
509	pet food manufacturers association uk (pfma)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
510	pet food processing	SGS Digicomply Dataset	SGS Digicomply
511	petfoodindustry.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
512	pfas central	SGS Digicomply Dataset	SGS Digicomply

513	philippine bureau of agriculture and fisheries standards (bafs)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
514	plastic europe	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
515	plastic soup foundation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
516	plos blogs	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
517	polish federation of food producers union of employers (pfpz)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
518	portafolio.co	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
519	portal.efet.gr	Food Recall Incidents	Agroknow
520	portman group	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
521	pr newswire	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
522	premium times	SGS Digicomply Dataset	SGS Digicomply
523	presidencia de la república oriental del uruguay	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
524	press information bureau	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
525	processed foods registration directorate	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
526	produce report	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
527	public health agency of canada	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
528	punto focal	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
529	quadram institute	SGS Digicomply Dataset	SGS Digicomply
530	que choisir	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
531	rapid alert system for food and feed portal (rasff)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
532	rappel conso	SGS Digicomply Dataset	SGS Digicomply
533	recalls-rappels.canada.ca	Food Recall Incidents, Food Recalls	Agroknow
534	regulatory focus	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
535	republic of china import refusals	Border Rejections	Agroknow
536	republic of the philippines - department of health	SGS Digicomply Dataset	SGS Digicomply
537	republic of the philippines - house of representatives	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
538	republic of turkey ministry of agriculture and forestry	Food Recalls	Agroknow
539	researchgate	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
540	rijksoverheid	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

541	rosstandart federal agency on technical regulating and metrology	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
542	russian center of agroanalytics	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
543	russian federation center of veterinary	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
544	samacharcentral.com	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
545	santé publique france	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
546	saudi food & drug authority	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
547	saudi food and drug authority	Food Recalls	Agroknow
548	scidevnet	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
549	science direct	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
550	science publications	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
551	scitechdaily	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
552	scotch whisky association	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
553	scottish government	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
554	seachoice	SGS Digicomply Dataset	SGS Digicomply
555	seafood watch	SGS Digicomply Dataset	SGS Digicomply
556	securite-alimentaire.public.lu	Food Recall Incidents	Agroknow
557	serious eats	SGS Digicomply Dataset	SGS Digicomply
558	service public fédéral justice	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
559	services.parliament.uk	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
560	servicio ecuatoriano de normalización	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
561	sgs digicomply editorial team	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
562	sgs sa	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
563	sinab - national information system on organic agriculture (italy)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
564	singapore food agency (former agri-food & veterinary authority of singapore)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
565	singapore national environment agency	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
566	singapore statutes online	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
567	sirim	SGS Digicomply Dataset	SGS Digicomply

568	sistem informasi direktorat jenderal peraturan perundang-undangan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
569	skal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
570	sky news	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
571	smithers	SGS Digicomply Dataset	SGS Digicomply
572	social market foundation	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
573	south africa government	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
574	south african bureau of standards (sabs)	SGS Digicomply Dataset	SGS Digicomply
575	south china morning post	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
576	south korean ministry of environment	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
577	southeast farm press	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
578	springerlink	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
579	sputnik	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
580	standartization, metrology and certification agency of uzbekistan «uzstandard»	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
581	state administration for market regulation	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
582	state of israel - ministry of health	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
583	state veterinary and food administration - svps sr - slovakia	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
584	sustainable packaging coalition	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
585	swedish nfa - livsmedelsverket	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
586	switzerland federal council	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
587	t.c. gıda tarım ve hayvancılık bakanlığı - republic of turkey ministry of food, agriculture and livestock	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
588	taipei times	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
589	taiwan environmental protection administration	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
590	taiwan fda food drug consumer knowledge service network	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
591	taiwan food and drug administration (fda)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
592	taylor & francis online	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
593	technology.org	SGS Digicomply Dataset	SGS Digicomply

594	tegengif	SGS Digicomply Dataset	SGS Digicomply
595	thai government gazette	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
596	the alcohol beverages advertising code scheme	SGS Digicomply Dataset	SGS Digicomply
597	the american association for agriculture education (aaae)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
598	the australian wine research institute	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
599	the british specialist nutrition association limited ("bsna")	SGS Digicomply Dataset	SGS Digicomply
600	the china food law blog	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
601	the conversation us	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
602	the daily intake	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
603	the european consumer organisation (beuc)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
604	the executive yuan gazette online (taiwan)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
605	the fish site	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
606	the florida senate	SGS Digicomply Dataset	SGS Digicomply
607	the food and beverage news	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
608	the food tech	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
609	the german food association (der lebensmittelverband deutschland)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
610	the gulf standardization organization (gso)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
611	the hill	SGS Digicomply Dataset	SGS Digicomply
612	the japan food chemical research foundation	SGS Digicomply Dataset	SGS Digicomply
613	the kathmandu post	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
614	the legal publisher lexxion	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
615	the public authority for consumer protection - pacp - oman	SGS Digicomply Dataset	SGS Digicomply
616	the republic of serbia - ministry of agriculture and environmental protection - plan protection directorate	SGS Digicomply Dataset	SGS Digicomply
617	the spoon	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
618	the standards institution of israel	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
619	the wall street journal	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply



620	the washington post	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
621	third-party weather data provider	Weather data	Agrivi
622	tiempo.com.mx	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
623	tna.europarchive.org	Food Recall Incidents	Agroknow
624	toxic-free future	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
625	trade.ec.europa.eu	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
626	treehugger	SGS Digicomply Dataset	SGS Digicomply
627	trevisan & cuonzo law firm	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
628	trinidad & tobago bureau of standards	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
629	truthaboutpetfood.com	SGS Digicomply Dataset	SGS Digicomply
630	tuoitrenews.vn	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
631	turkey ministry of agriculture and forestry	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
632	turkish official gazette	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
633	u.s. department of agriculture	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
634	u.s. department of agriculture (usda)	Food Recalls	Agroknow
635	u.s. federal register	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
636	u.s. house committee on oversight and reform	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
637	unified site for publication of draft legal acts - armenia	SGS Digicomply Dataset	SGS Digicomply
638	united nations environment programme (unep)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
639	united nations system standing committee on nutrition	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
640	united press international (upi)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
641	united states department of agriculture - agricultural marketing service	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
642	united states department of agriculture - economic research service	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
643	united states environmental protection agency	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
644	university of nebraska–lincoln	SGS Digicomply Dataset	SGS Digicomply
645	us senator for new hampshire hassan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
646	usa regulations.gov	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
647	usa today	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
648	usda - food and nutrition service	SGS Digicomply Dataset	SGS Digicomply



649	usda - foreign agricultural service	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
650	usda foreign agricultural service - global agricultural information network	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
651	verein für konsumenteninformation (vki)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
652	verkhovna rada of ukraine	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
653	vermont law edu	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
654	vietnam association of seafood exporters and producers (vasep)	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
655	vietnam ministry of finance	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
656	visual crossing	Weather data	Agroknow
657	vrije universiteit amsterdam	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
658	văn phòng sps việt nam (vietnam sanitary and phytosanitary notification authority and enquiry point)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
659	wam en	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
660	washington state department of agriculture	Food Recalls	Agroknow
661	waste dive	Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
662	wayback.archive-it.org	Food Recall Incidents	Agroknow
663	webarchive.nationalarchives.gov.uk	Food Recall Incidents	Agroknow
664	welsh government	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
665	wiley online library - open content	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
666	world economic forum	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
667	world health organization (who)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
668	world trade organization (wto)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
669	wri.org	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
670	wto documents online (notifications)	FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset	SGS Digicomply
671	www.collectionscanada.gc.ca	Food Recall Incidents	Agroknow
672	www.food.gov.uk	Food Recall Incidents	Agroknow
673	www.fsai.ie	Food Recall Incidents	Agroknow
674	www.interpol.int	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
675	www.salute.gov.it	Food Recall Incidents	Agroknow

676	yumda.de	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
677	zero waste europe	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
678	zerosette.it	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
679	énfasis alimentación	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
680	ökotest	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
681	россельхознадзор - rosselkhoznadzor - federal service for veterinary and phytosanitary surveillance	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
682	القاهرة 24	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
683	صحيفة أثير atheer oman	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
684	식품저널 인터넷식품신문 - food journal internet food newspaper	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
685	Mtech Platform	Poultry Production Data	Agroknow
686	Food Fortress Platform	Food Fortress Dataset	Agroknow

**Table 44:** The languages supported by the datasets provided by the owners to support KPI 1.11

ID	Language Code	Language Name	Datasets	Owners
1	ar	Arabic	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
2	ast	Asturian	SGS Digicomply Dataset	SGS Digicomply
3	az	Azerbaijani	SGS Digicomply Dataset	SGS Digicomply
4	bg	Bulgarian	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
5	bn	Bangla	SGS Digicomply Dataset	SGS Digicomply
6	br	Breton	SGS Digicomply Dataset	SGS Digicomply
7	bs	Bosnian	SGS Digicomply Dataset	SGS Digicomply
8	ca	Catalan	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
9	cs	Czech	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
10	cy	Welsh	SGS Digicomply Dataset	SGS Digicomply
11	da	Danish	Food Recall Incidents, Food Recalls, FoodSafeSum, SGS Digicomply Dataset	Agroknow, SGS Digicomply
12	de	German	Food Recall Incidents, FoodSafeSum, SGS Digicomply Dataset, Food Recalls	Agroknow, SGS Digicomply
13	el	Greek	Food Recall Incidents, Food Recalls, FoodSafeSum, SGS Digicomply Dataset	Agroknow, SGS Digicomply

14	en	English	Food Recall Incidents, FoodSafeSum, Chunk Relevance Validation Dataset, SGS Digicomply Dataset, CoSRec, Border Rejections, Food Recalls, Foodborne Disease Outbreaks, Inspections, Food Safety News, Abstracts of a systematic review on Food and federated learning, Abstracts of a systematic review on Food and NLP, News Articles by WFSR	SGS Digicomply, CNR, Agroknow, WFSR
15	es	Spanish	FoodSafeSum, SGS Digicomply Dataset, Border Rejections, Food Recalls	SGS Digicomply, Agroknow
16	et	Estonian	SGS Digicomply Dataset	SGS Digicomply
17	fa	Persian	SGS Digicomply Dataset	SGS Digicomply
18	fi	Finnish	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
19	fr	French	Food Recall Incidents, FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
20	gu	Gujarati	SGS Digicomply Dataset	SGS Digicomply
21	he	Hebrew	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
22	hi	Hindi	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
23	hr	Croatian	SGS Digicomply Dataset	SGS Digicomply
24	hu	Hungarian	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
25	hy	Armenian	SGS Digicomply Dataset	SGS Digicomply
26	id	Indonesian	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
27	ilo	Iloko	SGS Digicomply Dataset	SGS Digicomply
28	is	Icelandic	SGS Digicomply Dataset	SGS Digicomply
29	it	Italian	Food Recall Incidents, FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
30	iw	Hebrew	SGS Digicomply Dataset	SGS Digicomply
31	ja	Japanese	FoodSafeSum, SGS Digicomply Dataset, Border Rejections	SGS Digicomply, Agroknow
32	km	Khmer	SGS Digicomply Dataset	SGS Digicomply
33	ko	Korean	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
34	lt	Lithuanian	SGS Digicomply Dataset	SGS Digicomply
35	lv	Latvian	SGS Digicomply Dataset	SGS Digicomply
36	mk	Macedonian	FoodSafeSum	SGS Digicomply

37	ms	Malay	SGS Digicomply Dataset	SGS Digicomply
38	mt	Maltese	SGS Digicomply Dataset	SGS Digicomply
39	ne	Nepali	SGS Digicomply Dataset	SGS Digicomply
40	nl	Dutch	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
41	no	Norwegian	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
42	pa	Punjabi	SGS Digicomply Dataset	SGS Digicomply
43	pl	Polish	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
44	pt	Portuguese	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
45	ro	Romanian	SGS Digicomply Dataset	SGS Digicomply
46	ru	Russian	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
47	sk	Slovak	SGS Digicomply Dataset	SGS Digicomply
48	sl	Slovenian	SGS Digicomply Dataset	SGS Digicomply
49	sm	Samoan	SGS Digicomply Dataset	SGS Digicomply
50	so	Somali	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
51	sr	Serbian	SGS Digicomply Dataset	SGS Digicomply
52	sv	Swedish	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
53	th	Thai	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
54	tl	Tagalog	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
55	tr	Turkish	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
56	uk	Ukrainian	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply
57	und	Unknown language	SGS Digicomply Dataset	SGS Digicomply
58	uz	Uzbek	SGS Digicomply Dataset	SGS Digicomply
59	vi	Vietnamese	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
60	zh	Chinese	SGS Digicomply Dataset, Border Rejections	SGS Digicomply, Agroknow
61	zh-cn	Chinese (China)	FoodSafeSum, SGS Digicomply Dataset, Food Recalls	SGS Digicomply, Agroknow
62	zh-tw	Chinese (Taiwan)	FoodSafeSum, SGS Digicomply Dataset	SGS Digicomply

## Annex II

The structure of the data registry, including all fields used to capture dataset attributes, KPIs, and multilingual information, is illustrated in

A		B		C		D		E		F		G	
EFRA DATA REGISTRY 2025		General Information										Data Type (contained)	
DATA SOURCE		Description		Format	Language	Data Coverage		Size					
Details		A brief description of the data source		PDF, HTML, CSV, JSON, XML, etc	EN, FR, etc	The geographical coverage of the data (Region / Countries)		The size of the dataset (# of records)		Predicted values			
www.fda.gov		The Food Recall Incidents dataset consists of 7,546 short texts (from 5 to 360 characters each), which are the titles of food recall announcements (therefore referred to as title), and their full-length texts (from 56 to 48,318 characters). The data are crawled from 24 public food safety authority websites by Agroknow. Experts manually classified each text to four groups of classes describing hazards and products on two levels of granularity		csv	en	USA		1740		Food Product Recalls			
H		I		J		K		L		M		N	
Ownership & Updates		Accessibility											
Data Owner	URL / Access Point	License / Usage Rights	Access Restrictions	Data Quality/Accuracy	Data Provider within the Consortium	3rd Party Data Provider	Public / Private data source	Storage location	Related to EFRA Use Cases	Month of availability	Other Details		
Organization responsible for the data	The web address / location where the data can be accessed.	Information about the license and usage rights associated with the data.	Any access restrictions or requirements for using the data.	Any notes on the quality or accuracy of the data.	Partner who is responsible for aggregating the data source	External organisation responsible for aggregating the data source	Where the data is stored (e.g., specific server, cloud storage, database, EFRA Data Hub)	UC#1: Risk predictions for poultry pathogens, UC#2: Food Safety Optimal Pesticides Use, UC#3: Informing Regulatory Decisions with Food Risk Intelligence	Month in which the data source is available through the EFRA Data Hub				
Agroknow	https://doi.org/10.5281/zenodo.10820657	CC BY-NC-SA 4.0		manual annotation	Agroknow		public	Zenodo	UC#1				
										Peer-reviewed article			

**Figure 1.** This figure provides a snapshot of the key fields and example entries, thereby demonstrating how partner contributions and KPIs were systematically recorded and monitored.

A		B		C		D		E		F		G	
EFRA DATA REGISTRY 2025		General Information										Data Type (contained)	
DATA SOURCE		Description		Format	Language	Data Coverage		Size					
Details		A brief description of the data source			PDF, HTML, CSV, JSON, XML, etc	EN, FR, etc	The geographical coverage of the data (Region / Countries)		The size of the dataset (# of records)			Predicted values	
		The Food Recall Incidents dataset consists of 7,546 short texts (from 5 to 360 characters each), which are the titles of food recall announcements (therefore referred to as title), and their full-length texts (from 56 to 48,318 characters). The data are crawled from 24 public food safety authority websites by Agroknow. Experts manually classified each text to four groups of classes describing hazards and products on two levels of granularity											
www.fda.gov					csv	en	USA		1740			Food Product Recalls	
H		I		J		K		L		M		N	
O		P		Q		R		S		T		U	
Ownership & Updates					Accessibility								
Data Owner	URL / Access Point	License / Usage Rights	Access Restrictions	Data Quality/Accuracy	Data Provider within the Consortium	3rd Party Data Provider	Public / Private data source	Storage location	Related to EFRA Use Cases	Month of availability	Other Details		
Organization responsible for the data	The web address / location where the data can be accessed.	Information about the license and usage rights associated with the data.	Any access restrictions or requirements for using the data.	Any notes on the quality or accuracy of the data.	Partner who is responsible for aggregating the data source	External organisation responsible for aggregating the data source		Where the data is stored (e.g., specific server, cloud storage, database, EFRA Data Hub)	UC#1: Risk predictions for poultry pathogens, UC#2: Food Safety Optimal Pesticides Use, UC#3: Informing Regulatory Decisions with Food Risk Intelligence	Month in which the data source is available through the EFRA Data Hub			
Agroknow	<a href="https://doi.org/10.5281/zenodo.10820657">https://doi.org/10.5281/zenodo.10820657</a>	CC BY-NC-SA 4.0		manual annotation	Agroknow		public	Zenodo	UC#1				

**Figure 1:** Screenshots of the EFRA Data Registry used for monitoring data population