



Co-funded by
the European Union



Spatio-TEmporal Linked data tools for the AgRI-food data space

Project Overview

Data Week 2023
Lulea

Manolis Koubarakis
National and Kapodistrian University of Athens

stelar-project.eu



Project Factsheet

Title: STELAR: Spatio-TEmporal Linked data tools for the AgRi-food data space

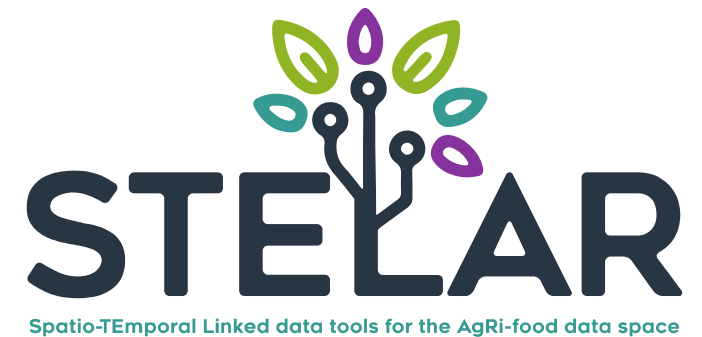
Topic: HORIZON-CL4-2021-DATA-01-03: Technologies for data management

Type of action: HORIZON Innovation Actions

Project starting date: 1 September 2022

Project duration: 36 months

Grant number: 101070122



<https://stelar-project.eu/>



Objectives



Provide tools to improve data discovery, data integration and AI-readiness of data.





Integrate these tools into an innovative Knowledge Lake Management System (KLMS).



Evaluate the KLMS in real-world data management challenges in the agrifood sector.

Motivation

Data Lakes

-  Large amounts of heterogeneous data in their original form, allowing data scientists to perform ad hoc, self-service analytics.
-  Difficult to discover relevant data, to integrate data from different sources, and to ensure that data is of high quality.

Agri-food Data Space

-  One of the main data spaces according to the European Strategy for Data due to its high importance for health, the economy, and the environment.
-  The agri-food sector is witnessing increasing digitalization and transformation; however, it is still lagging behind.

Approach

STELAR will develop a **Knowledge Lake Management System (KLMS)** to support **FAIR** and **AI-ready** data, (semi-)automatically turning a raw data lake into a knowledge lake. This is achieved by:

Enhancing the data lake with a **knowledge layer**:

- a Data Catalog offering automatically enhanced metadata for the data assets in the lake
- a Knowledge Graph that semantically describes and interlinks these data assets using suitable domain ontologies and vocabularies

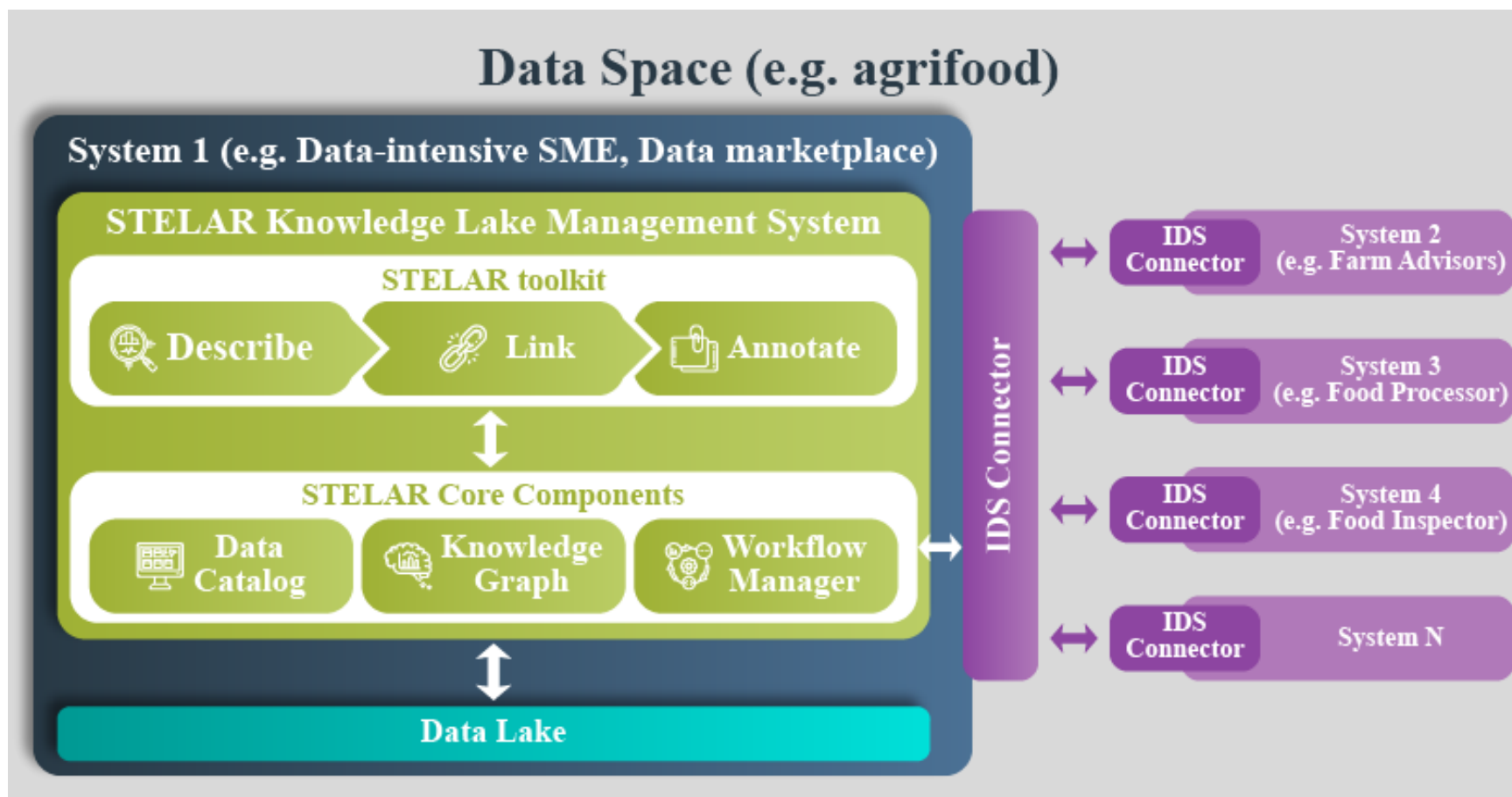
A set of **data management tools and workflows** for:

- data profiling, search and exploration
- textual, spatial and temporal data linking and alignment
- data labeling, data augmentation, bias detection

The **KLMS** will:

- enable **finding and selecting** relevant data assets
- facilitate **configuring and tuning** the data management tools
- support **designing, executing and monitoring** end-to-end data processing workflows

The STELAR KLMS



Pilot A: Risk Prevention in Food Supply Lines

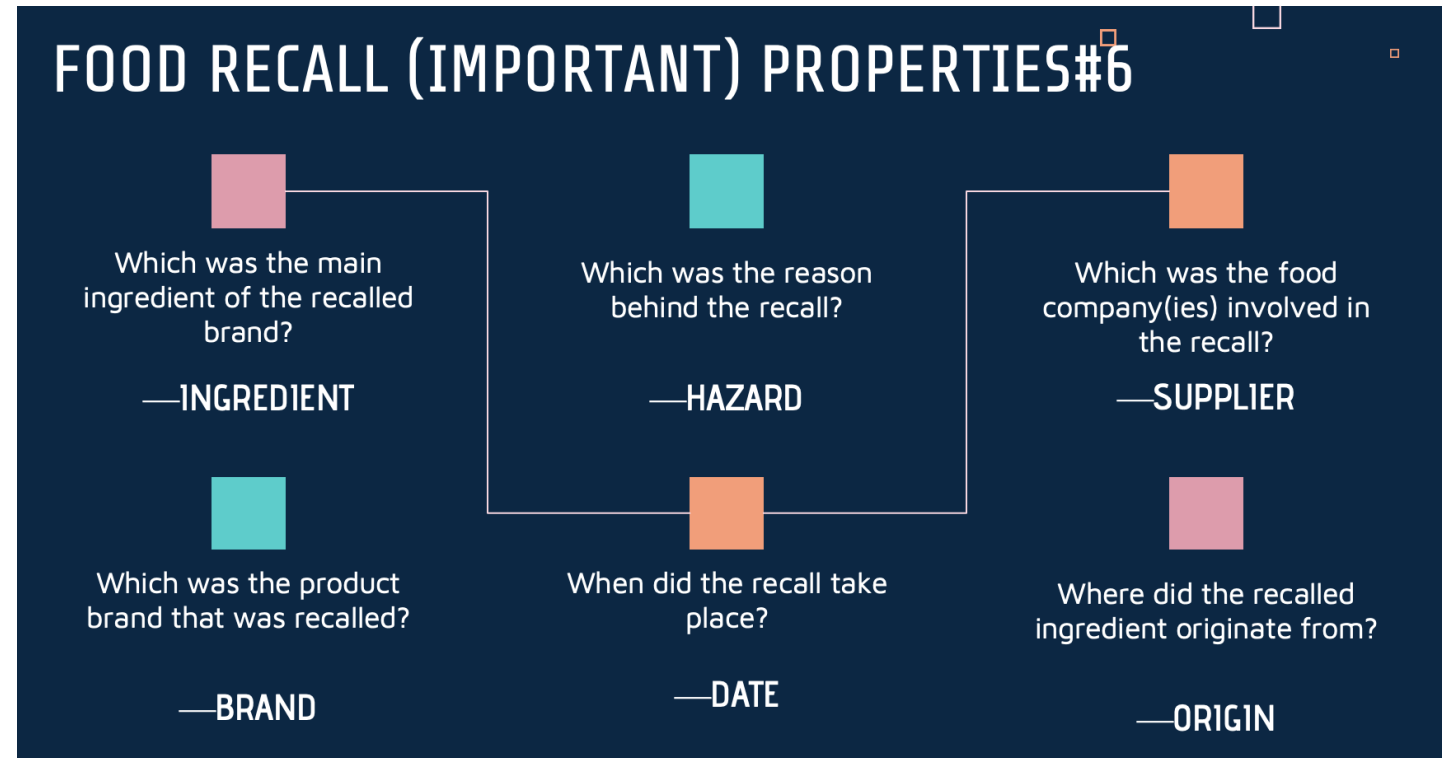


Data assets & vocabularies

- food incidents, inspections, lab tests, suppliers, prices, country indicators, news
- domain specific taxonomies for hazards and product classifications

Challenges

- frequent changes in data schema and format in the original sources
- human effort required to annotate the data records using terms from standard vocabularies
- alignment of different vocabularies used to classify products and hazards
- deduplication of food safety records published by different national authorities
- extracting location information for the food companies



Pilot B: Early Crop Growth Predictions

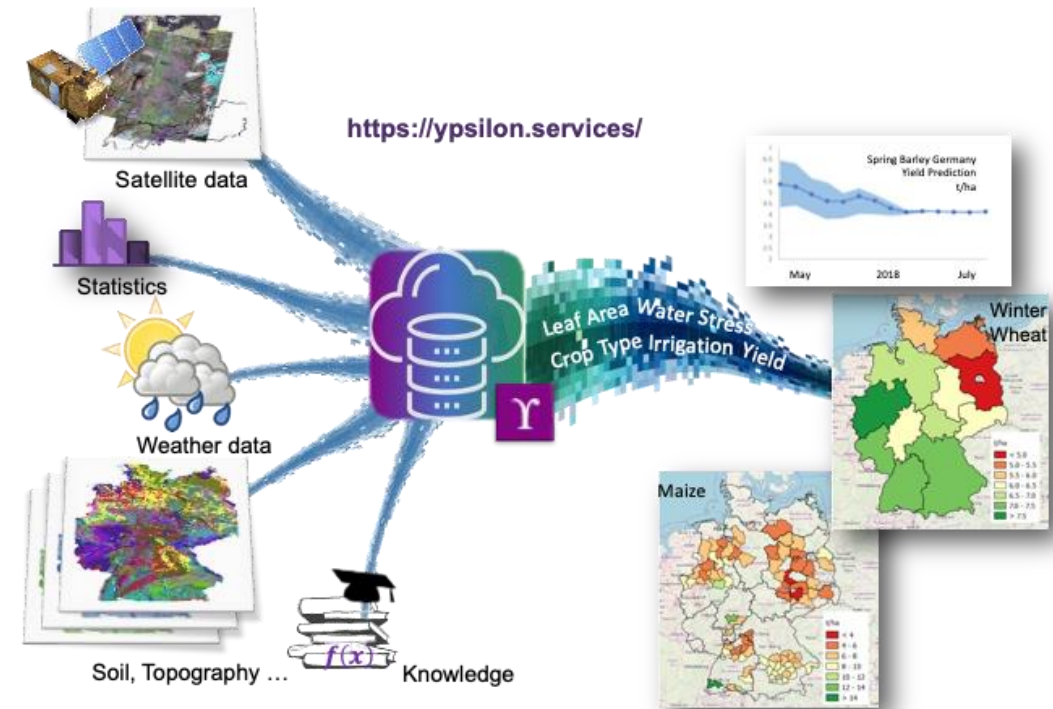


Data assets & vocabularies

- satellite data, meteorological data, seasonal forecasts, crop growth models
- vocabularies for plant development stages, crop types, and meteorological variables

Challenges

- data heterogeneity:
 - types of data (meteorological/satellite, measured/modeled)
 - spectral ranges (radar/optical)
 - spectral resolution (multi/hyperspectral)
 - spatial and temporal resolutions
- data volume:
 - high cost for data transfer and processing
- labeling:
 - annotated data from past years are needed for development, testing and validation
- processing:
 - data processing workflows need to be designed and configured



Pilot C: Timely Precision Farming Interventions



Data assets & vocabularies

- meteo-climatological data, soil data, crop plans, areas eligible for agricultural activity, assets and supply chain
- different vocabularies for product names, classifications and other metadata

Challenges

- aligning different product classifications
- combine and correlate data from heterogeneous sources to enable decision support





STELAR

Spatio-TEmporal Linked data tools for the AgRI-food data space

Thank you for your attention

Manolis Koubarakis
National and Kapodistrian University of Athens

koubarak@di.uoa.gr @mkoubarakis

stellar-project.eu

