CYBELE

Statistical Challenges Towards a Semantic Model for Precision Agriculture and Precision Livestock Farming

Dimitris Zeginis, **Evangelos Kalampokis,** Konstantinos Tarabanis

SemStats 2019

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 825355.

The CYBELE project

- Agriculture is a high volume, huge business with low operational efficiency
- Precision Agriculture and Livestock Farming use **intensive data collection and processing** to drive operational decisions
  - Drones patrol fields and alert farmers for crop ripeness or potential problems
  - Sensors on fields provide granular data points on soil conditions
  - GPS units on tractors can help determine optimal usage of heavy equipment
  - Satellite images can help computing useful field overview indicators e.g. Normalized Difference Vegetation Index
- The CYBELE project aims at demonstrating how Precision Agriculture and Livestock Farming can revolutionise the agrifood sector using the power of high performance computing

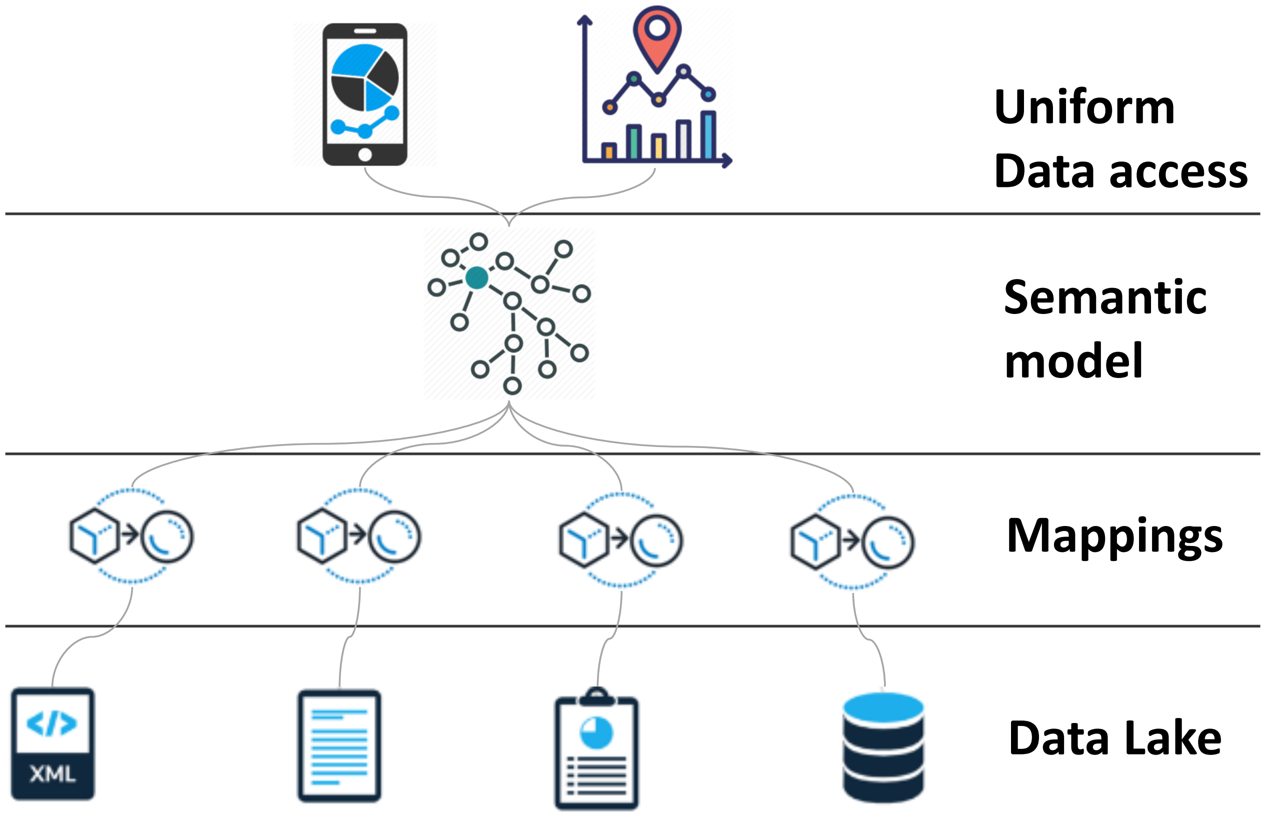
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Farming data

- Farming data come from diverse heterogeneous sources
- Structured data
  - Sensor data e.g. measure the soil's electrical conductivity at a specific location and time
  - Forecasts e.g. for weather, prices, production
- Unstructured data
  - Earth observations e.g. satellite/drone images
  - Video e.g. video data from pig pens to monitor pigs behaviour
  - Maps can be combined with other data to provide easily interpretable results
- Data lakes are required to store farming data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Uniform access to data lakes



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Role of the Semantic Model

- Represent domain knowledge related to the content of a data lake e.g. agriculture, farming, weather
- The semantic model can express:
  - Metadata:
    - Structural e.g. dimensions, measures
    - non-structural e.g. publisher, issuing date, license
  - Data:
    - values of dimensions e.g. geo dimension → Greece, New Zealand

V1 of the model

- Enables the uniform access of heterogeneous data
  - Facilitate data **discovery** → require metadata
  - Facilitate data **querying** → require data and metadata
  - Facilitate data **integration** → require data and metadata

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Semantic model development

- The methodology followed comprises the steps:
  - Study the scope of the model and the relevant data
  - Identify the user roles regarding data exploitation and their requirements
  - Extract the main concepts of the model from the requirements
  - Define the model by matching the concepts to existing standards and vocabularies

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Scope of the Semantic Model

- The semantic model focuses on the agri-food domain
  - Agriculture data e.g. protein content, soil electrical conductivity
  - Livestock farming data e.g. animal weight, livestock feed
  - Fishing data e.g. fish behavior data, landing data of fish stocks
  - Aquaculture data e.g. water temperature, current speed
  - Climate and weather data e.g. temperature, humidity
  - Satellite & aerial image data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

User roles

- End user (e.g. farmer and livestock manager)
  - exploit big data applications that produce easy to consume and interpret visualizations
- Modeler and developer
  - produce big data application &models for the end users
- Data analyst and farming consultant
  - exploit data-driven decision making to support end users
- Statistician
  - exploit big agricultural and livestock farming data to deliver official statistics

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Semantic Model User Requirements

- Search for datasets:
  - Regarding a specific cultivation e.g. soya, grapes
  - Created as a result of an activity e.g. sensoring, forecasting
  - That are updated e.g. monthly, daily, nearly real-time
  - Published/created/owned by a specific organization
  - Issued/modified after/before a specific point in time
  - That have a specific dimension e.g. geo, time
  - That have a specific measure e.g. NDVI
  - That have a specific unit of measure e.g. prices in euro
  - That have specific temporal coverage e.g. [2017- 2019]
  - Distributed in a specific format e.g. CSV, XML, JSON
  - Distributed under a specific license e.g. Creative Commons

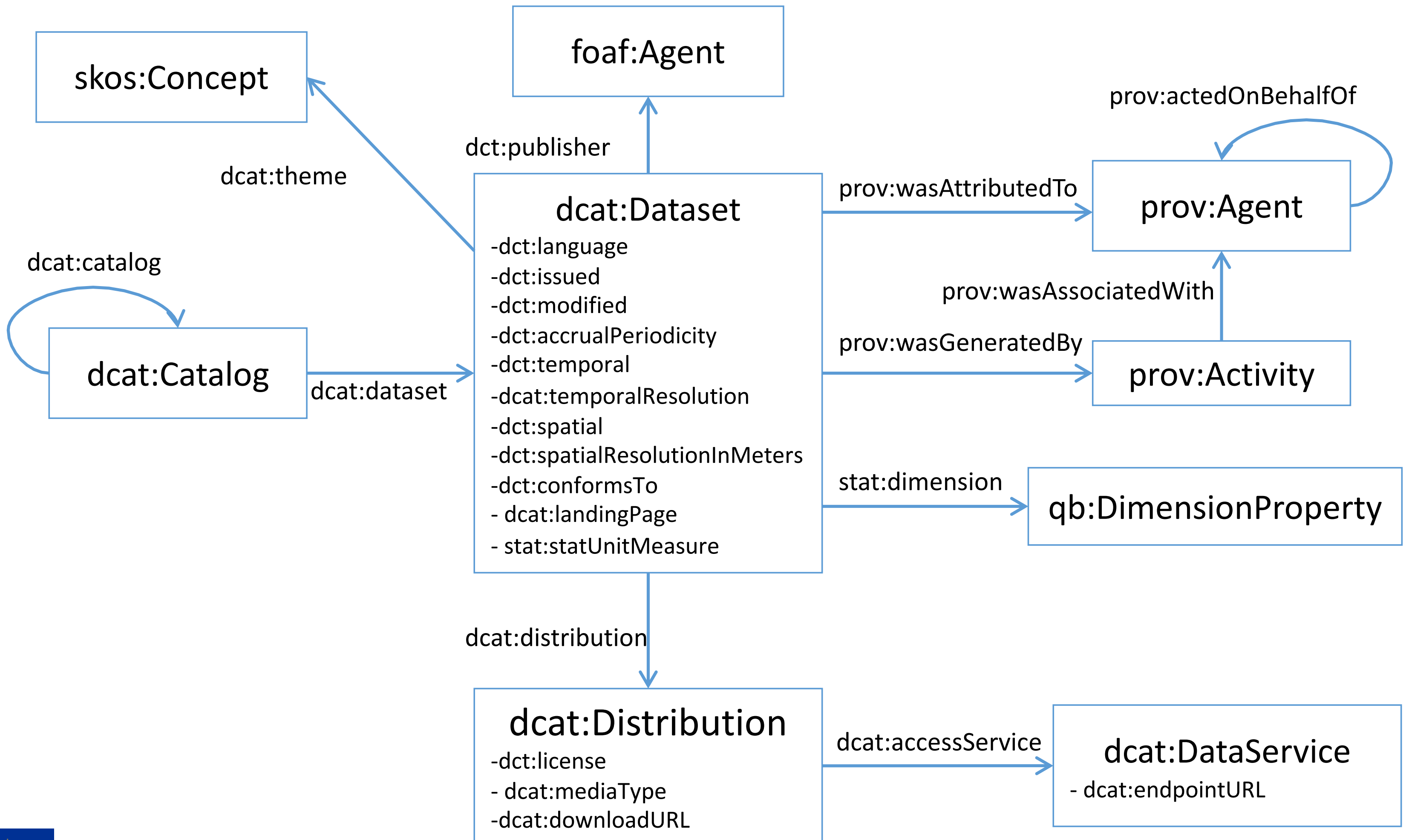
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Vocabularies used

- DCAT
  - describe datasets metadata
- Stat-DCAT
  - describe datasets statistical metadata
- PROV-O
  - describe provenance information
- QB vocabulary
  - describe statistical data and metadata

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

The model



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Statistical challenges

- Aggregated data are needed to support decision making
  - Sensors produce measurements regularly e.g. every 1 minute
  - Aggregated data are needed e.g. at day level
- Unstructured data need to be processed to calculate indexes
  - Satellites produce multispectral images
  - Indicators are needed e.g. Normalized Difference Vegetation Index (NDVI)
- Join of different datasets is required
  - Dataset 1: NDVI calculated from satellite images
  - Dataset 2: soil compression calculated from sensors at field
  - The join can use as an ID the field location

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Towards v2 of the model

Requirements:

- Requirement 1: query data
  - I want data of area X for the time [2018 - 2019] that measure the NDVI
  - Result: set of observations from one dataset
- Requirement 2: integrate data
  - I want data of area X for the time [2018 - 2019] that measure the NDVI AND the soil compression
  - Join observations from two datasets

Next steps:

- Define ontologies and code lists for:
  - Structural metadata: dimensions, measures, units
  - No-structural metadata: data format, theme, language, frequency update
  - Data values: time values, geo values, ...

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.

Thank you

https://www.cybele-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825355.