ARSINOE

CLIMATE RESILIENT-REGIONS THROUGH SYSTEMIC SOLUTIONS AND INNOVATIONS

SustainGraph: a Knowledge Graph for tracking the progress and the interlinking among the SDGs

Eleni Fotopoulou, Ioanna Mandilara, Christina-Maria Androna, Anastasios Zafeiropoulos, Symeon Papavassiliou

Institute of Communications and Computer Systems (ICCS), National Technical University of Athens (NTUA)

This project has received funding from the European Union's Horizon H2020 innovation action programme under grant agreement 101037424.

The SustainGraph Team

Eleni Fotopoulou
efotopoulou@netmode.ntua.gr

Ioanna Mandilara
ioannamand@netmode.ntua.gr

Christina-Maria Androna
androna xm@netmode.ntua.gr

Anastasios Zafeiropoulos
tzafeir@cn.ntua.gr

Prof. Symeon Papavassiliou
papavass@mail.ntua.gr

ΙCCS ΕΠΙΣΕΥ

NETMODE
NETWORK MANAGEMENT & OPTIMAL DESIGN LAB

https://www.netmode.ntua.gr/

ARSINOE

Network Management and Optimal Design Laboratory (NETMODE)

- The Institute of Communication and Computer Systems (ICCS) is a research institute of the School of Electrical and Computer Engineering (ECE) of the National Technical University of Athens (NTUA)
- Network Management and Optimal Design Laboratory (NETMODE)
- Consists of: 4 faculty members, 10 postdoctoral researchers, 20+ PhD students and research engineers
- NETMODE research concentrates on the design, optimization and management of heterogeneous networks and distributed dynamic systems, emphasizing on: optimization and orchestration in 5G/6G Networks, knowledge and data analysis, knowledge graphs, machine learning techniques, complex systems analysis, smart cities, and performance evaluation of stochastic systems.
- NETMODE members have extensive experience in Future Internet Research Experimentation and have deployed a Future Internet Testbed. NETMODE is actively involved in several RTD programs sponsored by National and European organizations.

https://www.netmode.ntua.gr/

ARSINOE

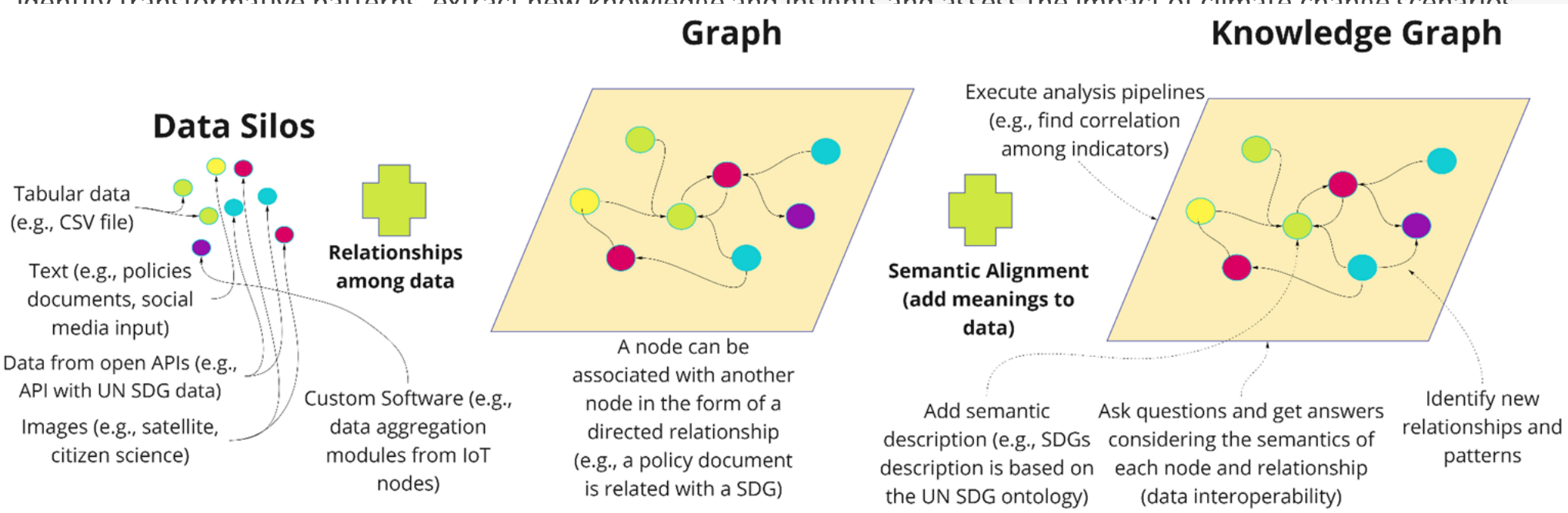
NETMODE
NETWORK MANAGEMENT & OPTIMAL DESIGN LAB

Why we need SustainGraph? (1/2)

- **Existence of many data silos around the SDGs**
  - SDG datasets in different formats and represented based on different semantics
  - Definition of indicators may differ per geographical area
  - Multiple policies documents and directives around the climate change
  - Customized software and Application Programming Interfaces (APIs) to do one thing
    - Data semantics are usually hidden from end users
  - Need for data interoperability and re-usability to develop resilient adaptation and mitigation solutions for the climate change
- **Development of SustainGraph**
  - a knowledge graph to represent data around the SDGs that can be interconnected and enriched with meaning to explicitly represent knowledge
  - Interlinking of the represented concepts with well-defined semantics
  - alignment of terminologies of the same concepts under different data schemas to facilitate interdisciplinary studies
  - data population mechanisms for time-series data, documents, files in tabular format, considering the time and spatial scale

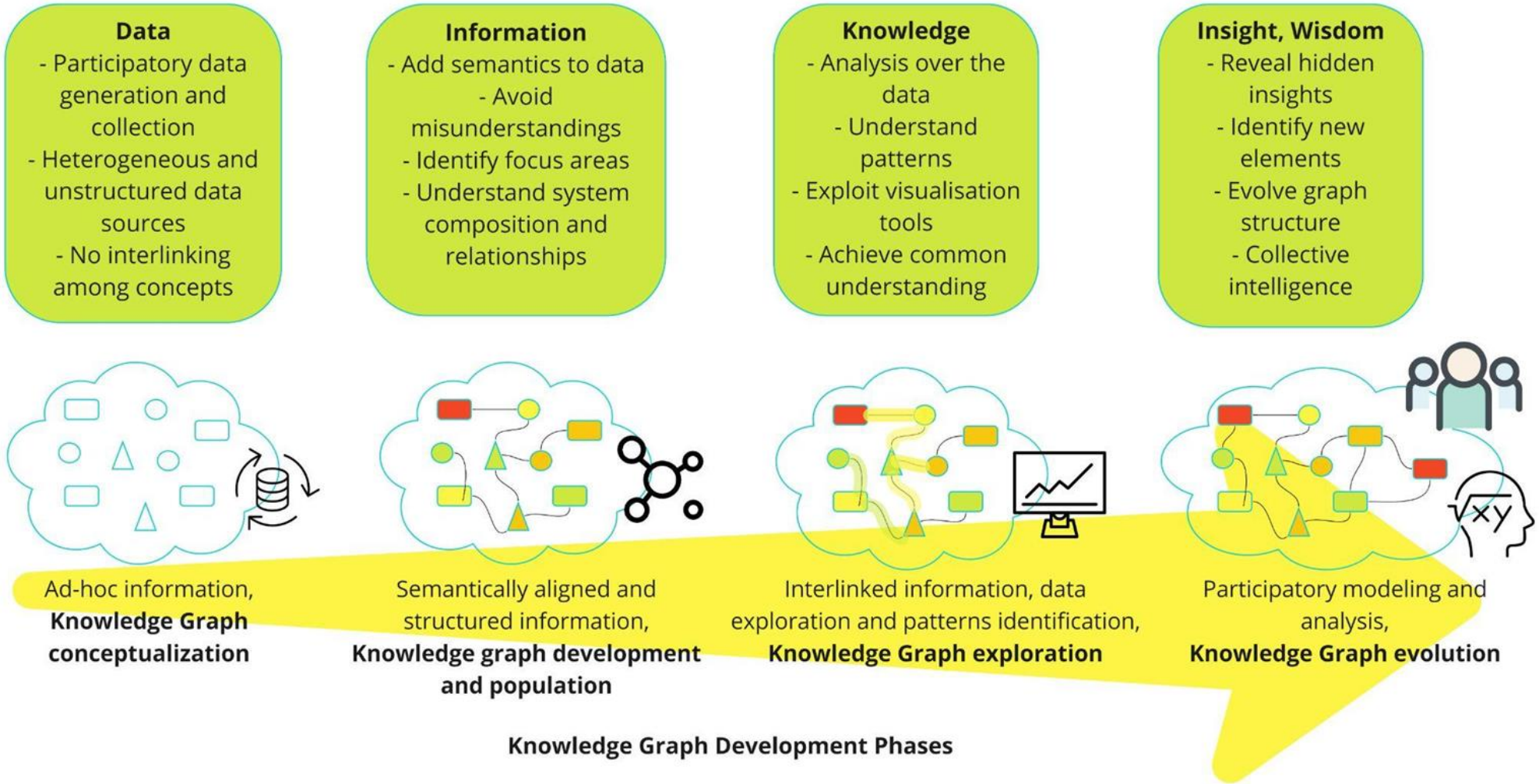
Why we need SustainGraph? (2/2)

- **Management of data volatility and assurance of data quality**
  - relationships among nodes can be dynamic
  - representation of complex and dynamic socio-ecological systems
  - quality management processes
- **Reasoning and analysis over the available data**
  - identification and prediction of new relationships among entities
  - participatory socio-environmental systems analysis
  - identify transformative patterns, extract new knowledge and insights and assess the impact of climate change scenarios



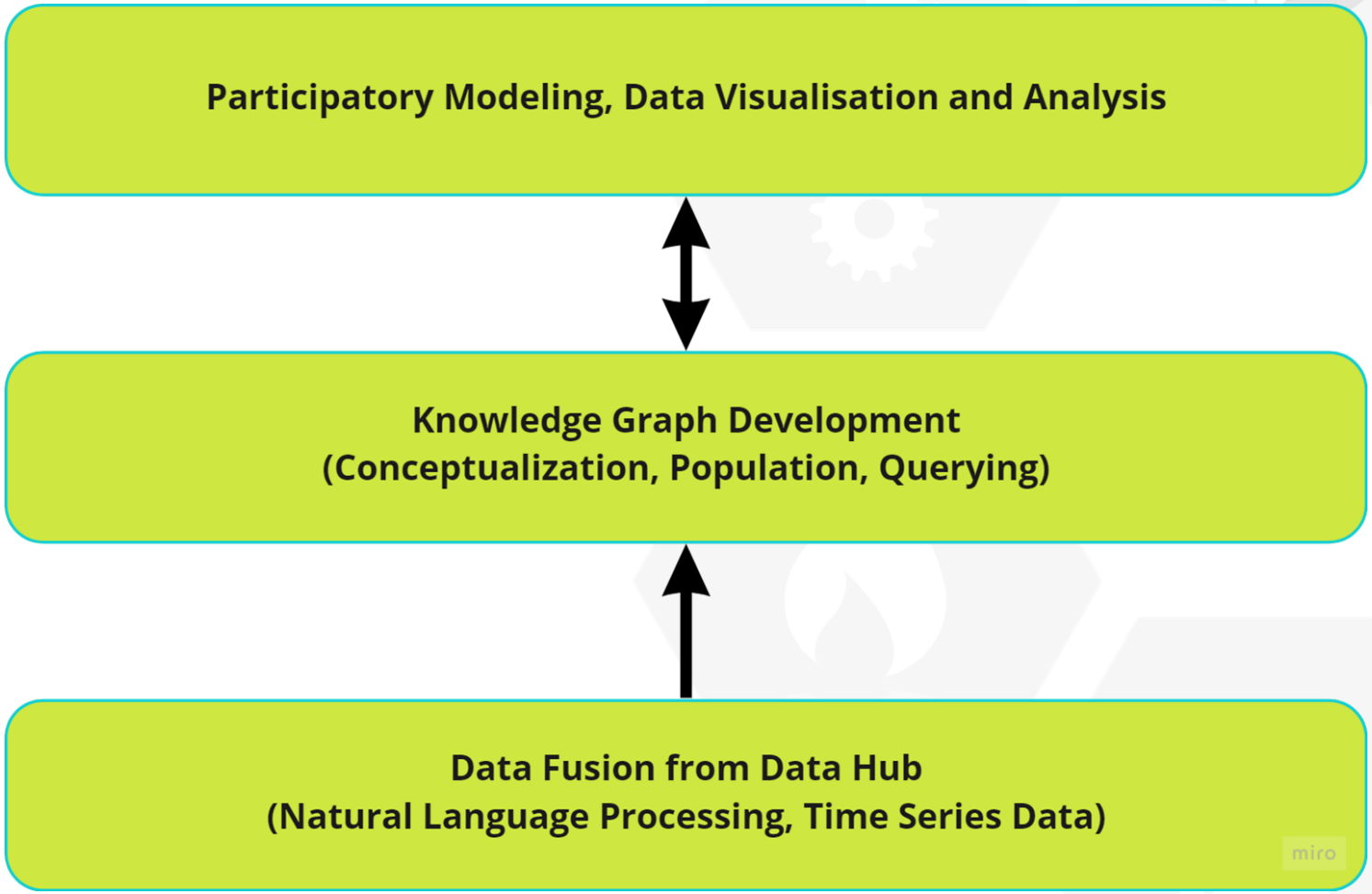
Interplay between a Systems Innovation Approach….

…. and a Knowledge Graph development

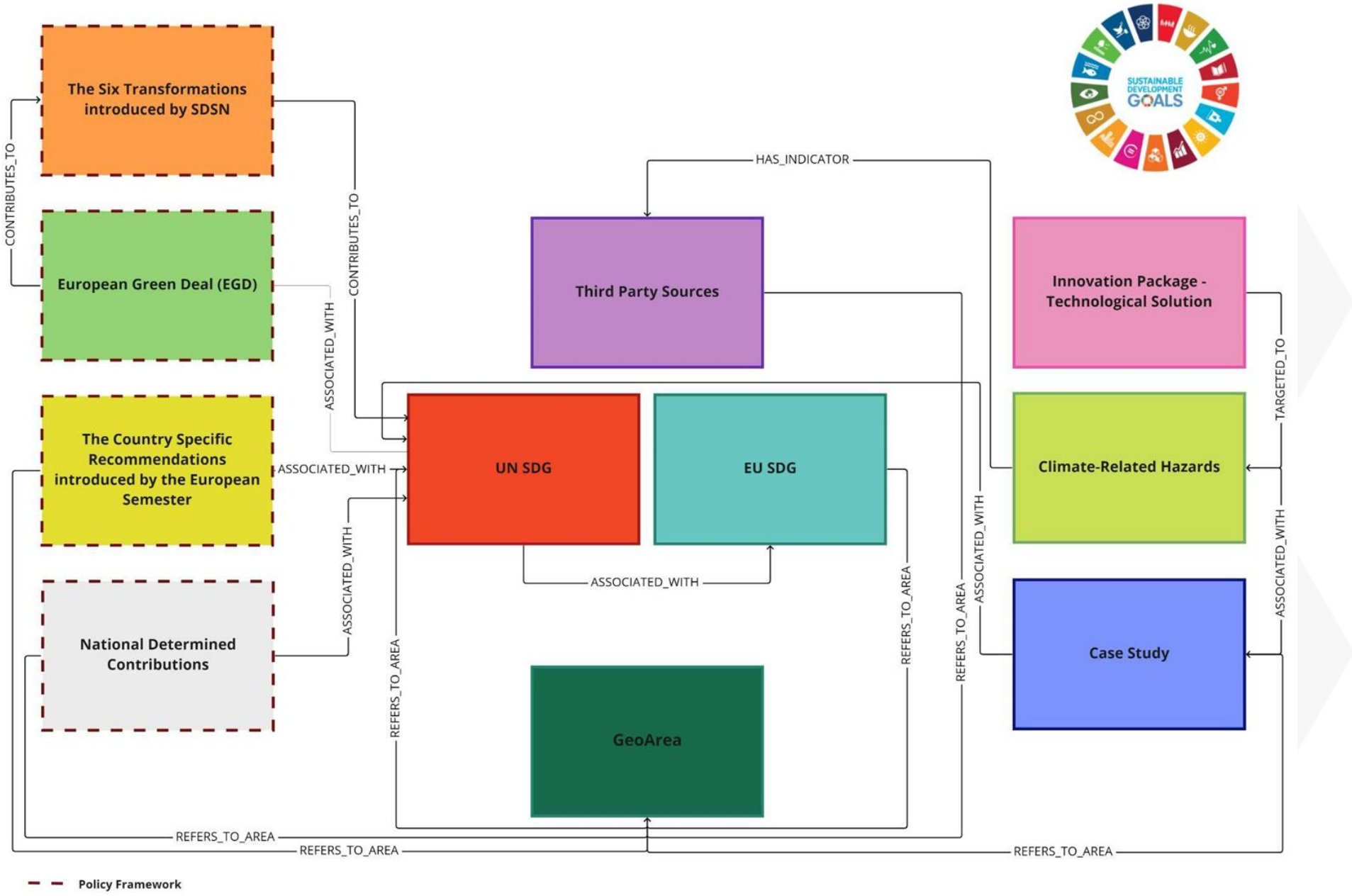


Knowledge Graph Ecosystem in ARSINOE

Layered Approach

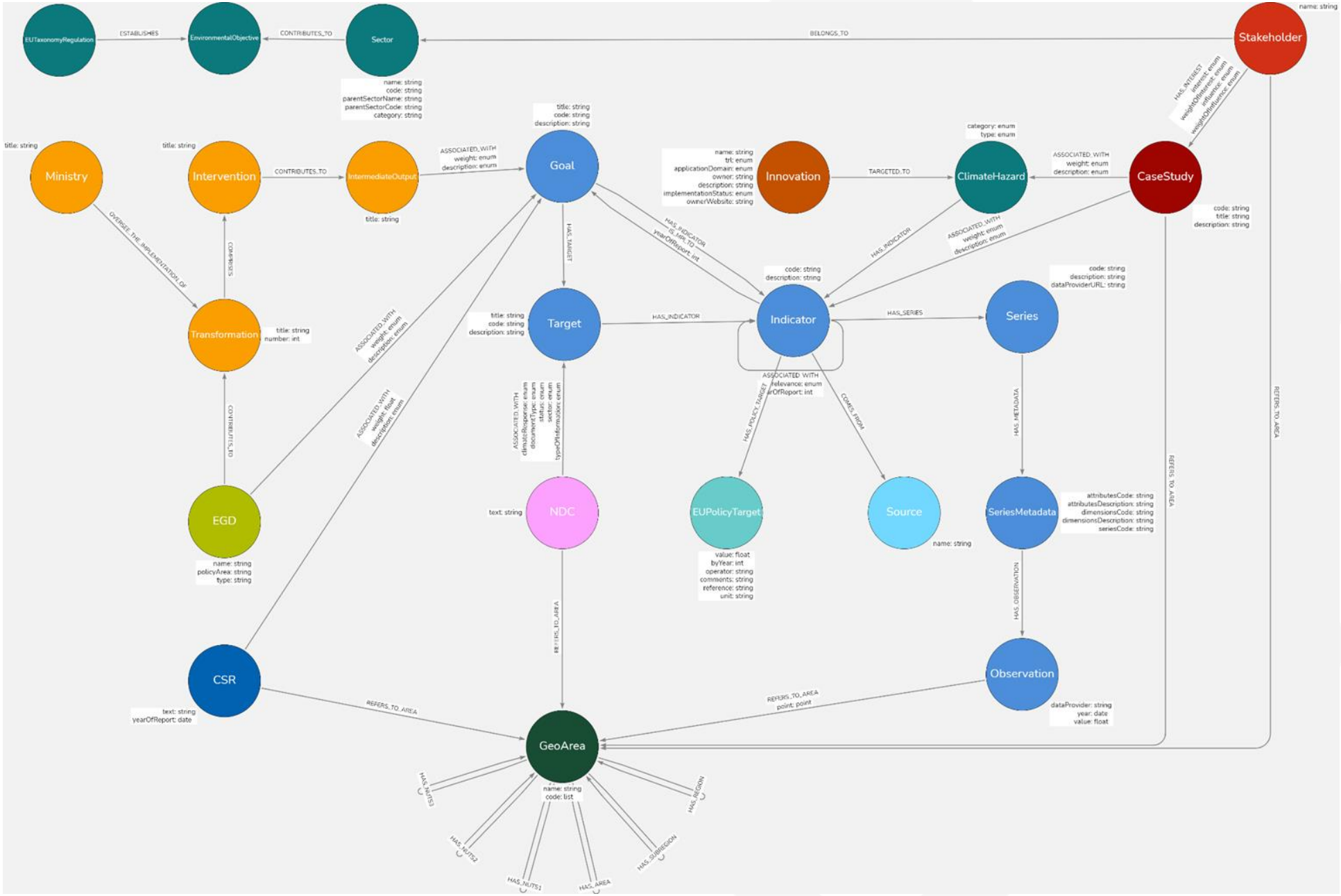


High level view of SustainGraph



[SustainGraph: A knowledge graph for tracking the progress and the interlinking among the sustainable development goals' targets]

Detailed view of SustainGraph



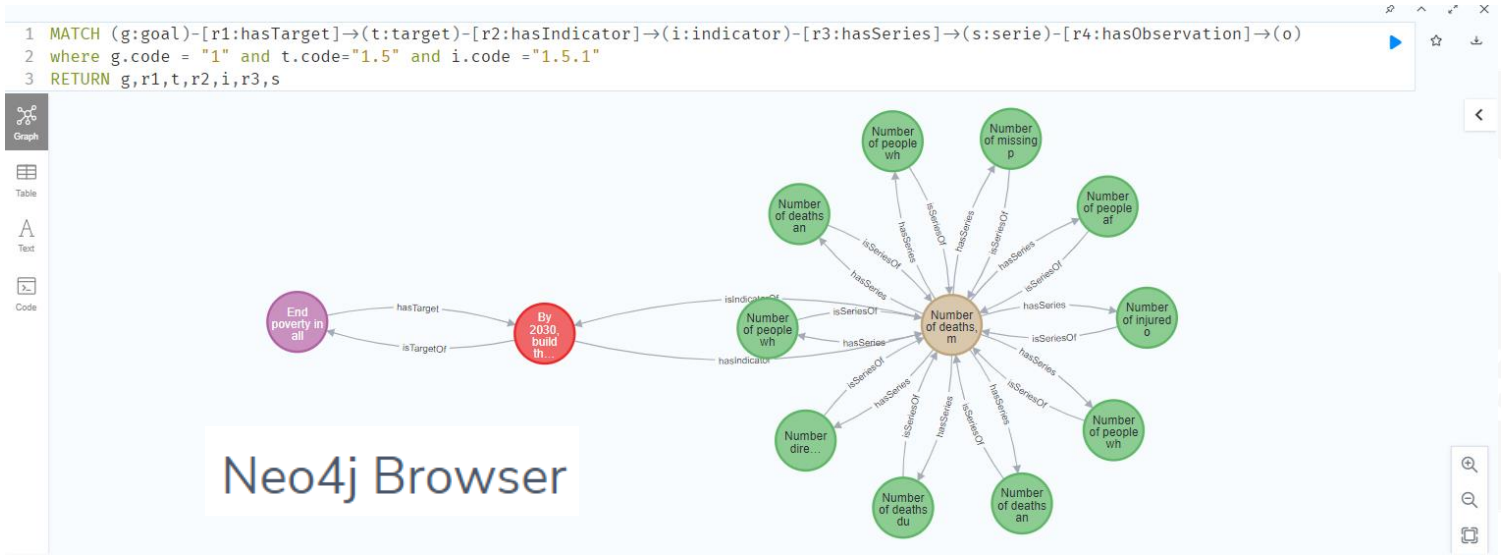
SustainGraph Key Characteristics

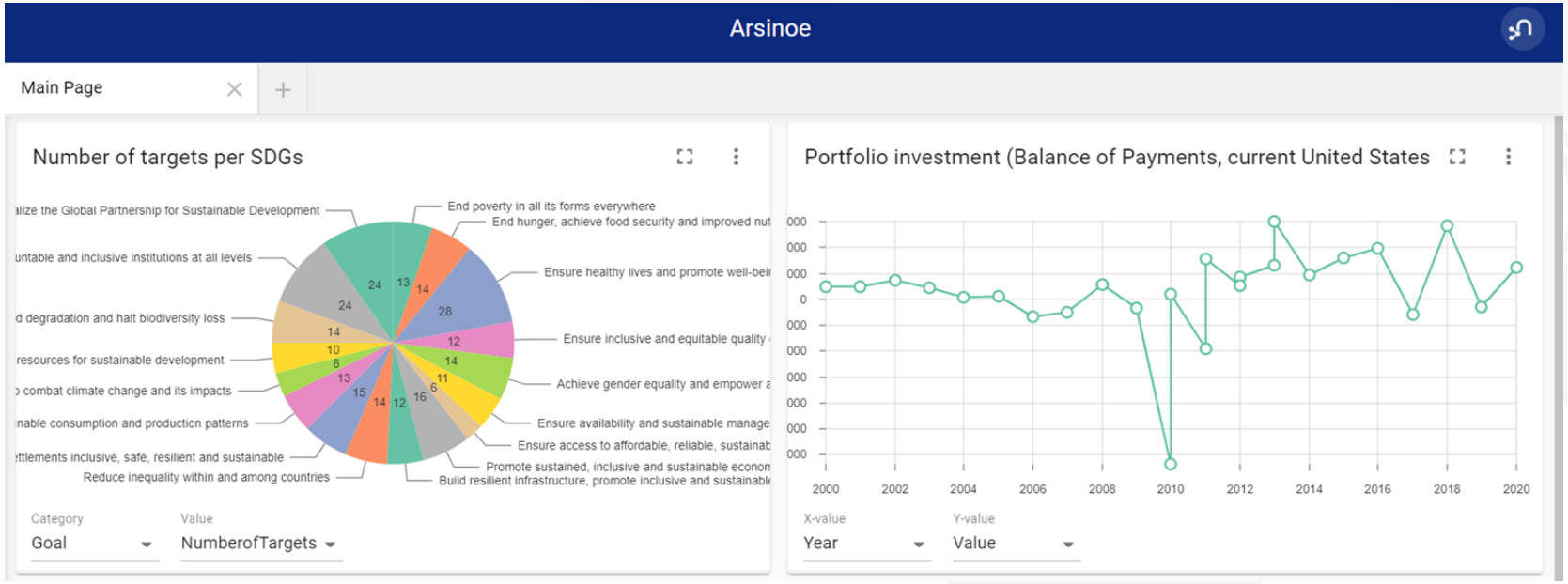
- Consider **various types** of data:
  - time series data
  - text/documents
  - tabular data
- Where applicable, include **geolocation** characteristics (e.g., geometry points from data coming from Copernicus service)
- Set of **data population pipelines** taking advantage of **ML techniques**
- Multiple **visualization tools** to cover needs of different stakeholders
- Alignment of terms with the **SustainGraph Ontology** (keem as much as possible semantic consistency and alignment)
  - Described using W3C RDF Schema and the Web Ontology Language.

https://gitlab.com/netmode/sustaingraph

https://netmode.gitlab.io/sustaingraph-ontology/

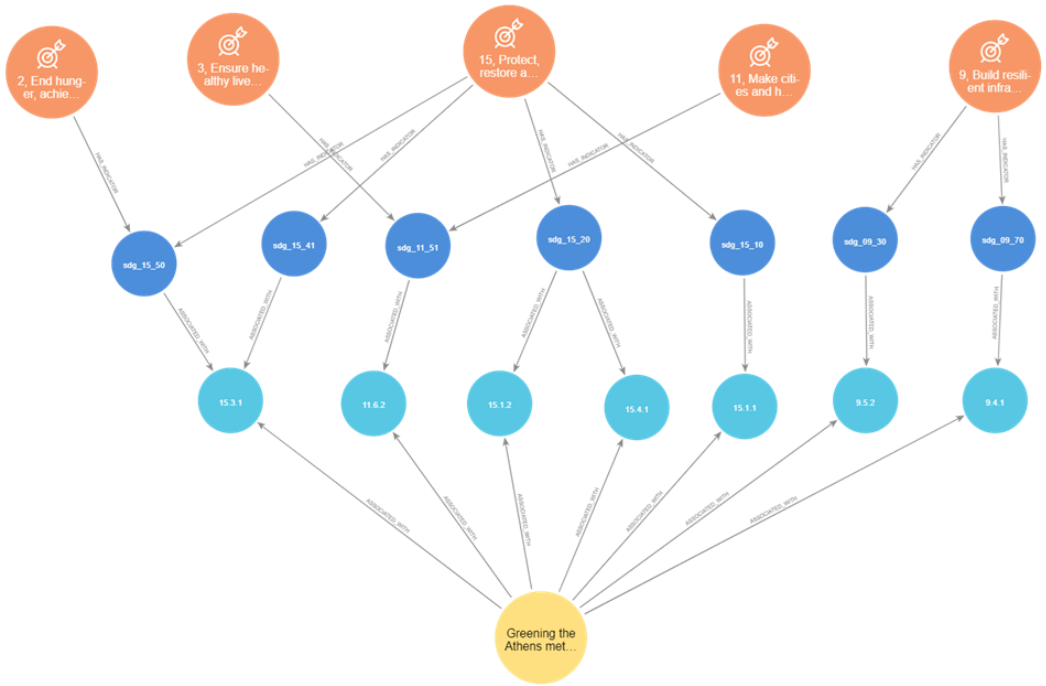
Querying and Visualization Tools (1/2)



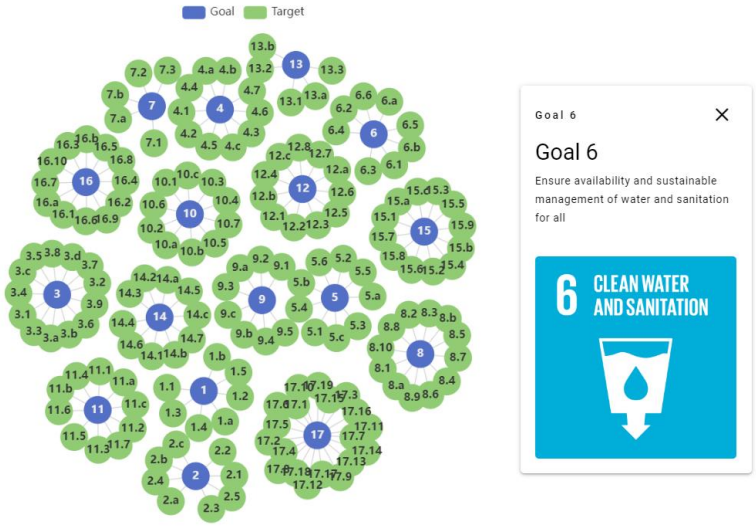


Querying and Visualization Tools (2/2)

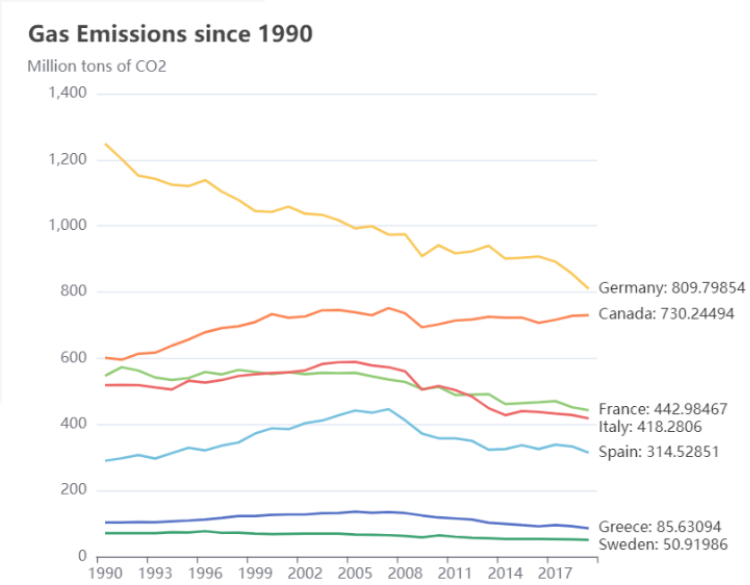
Neo4j Bloom



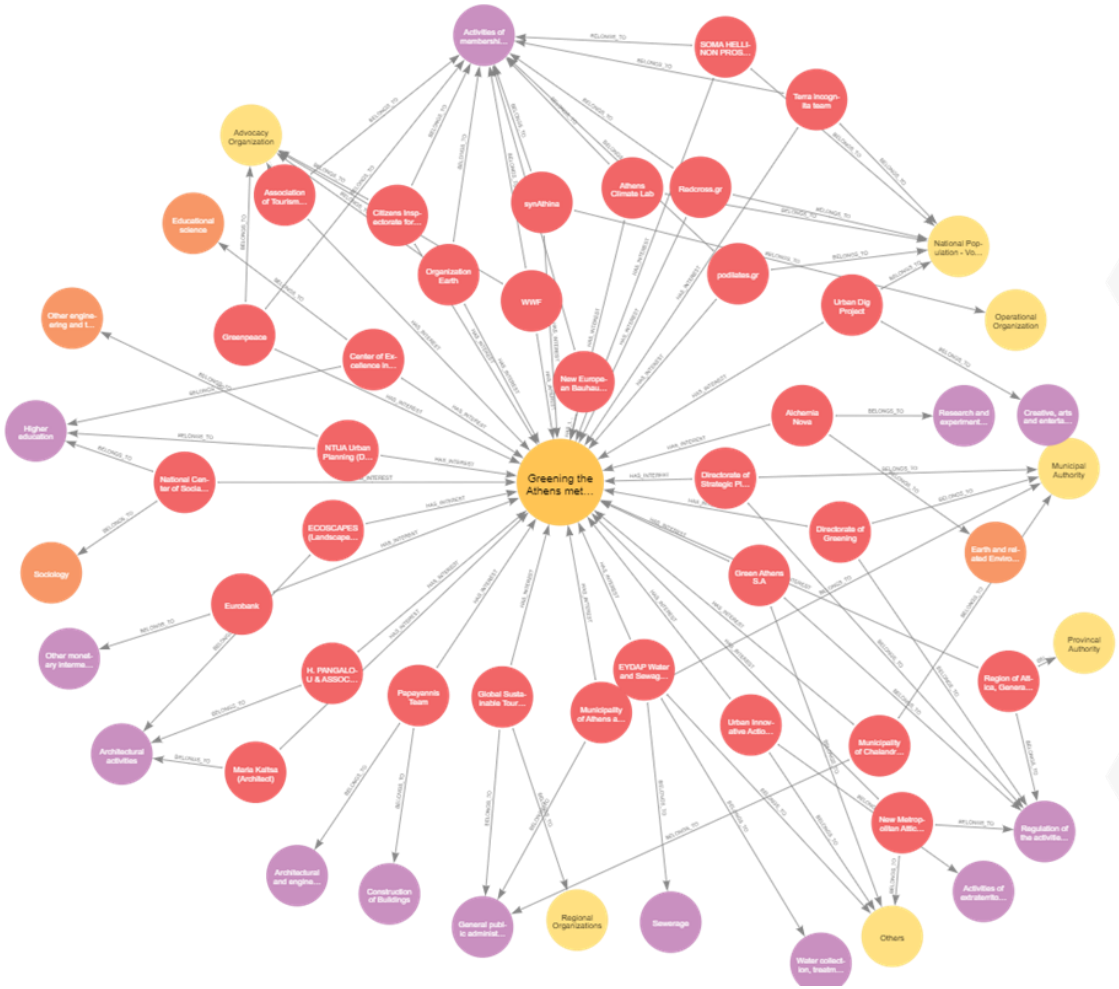
Relationship between the UN and the EU SDG indicators



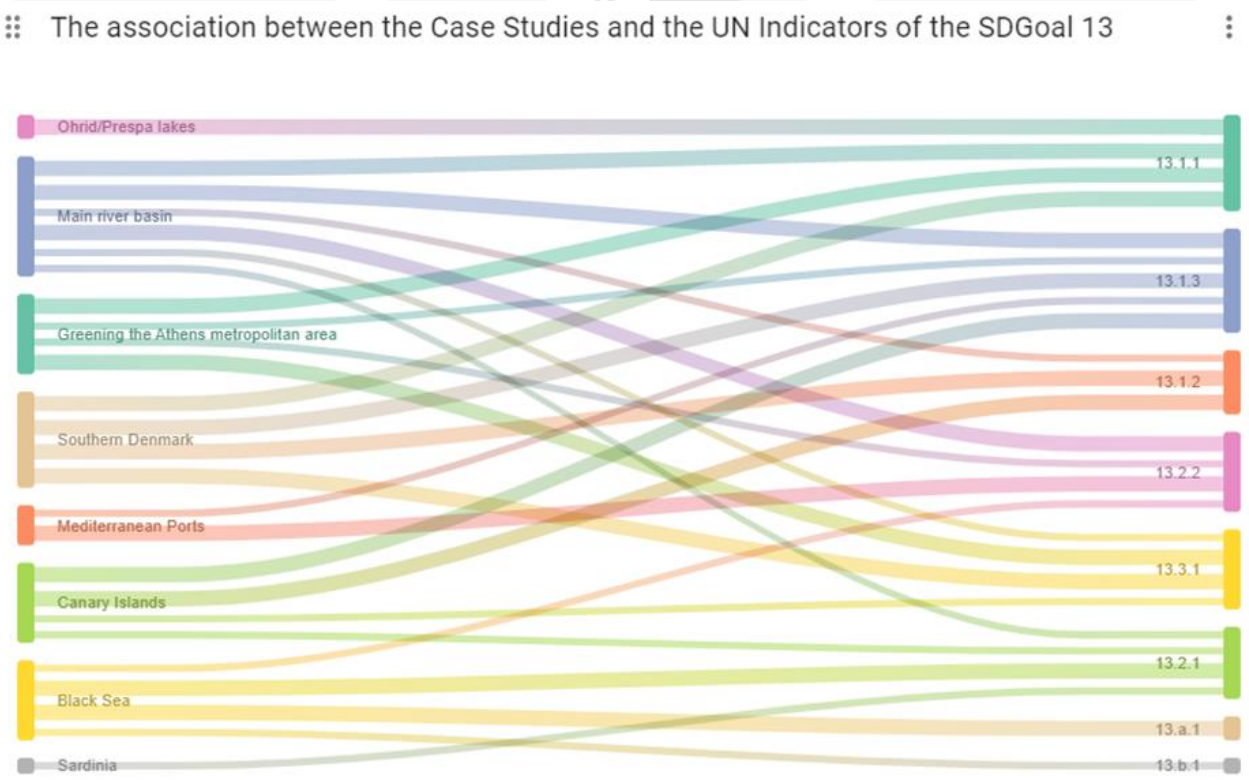
Custom Visualization kit based on Vue.js



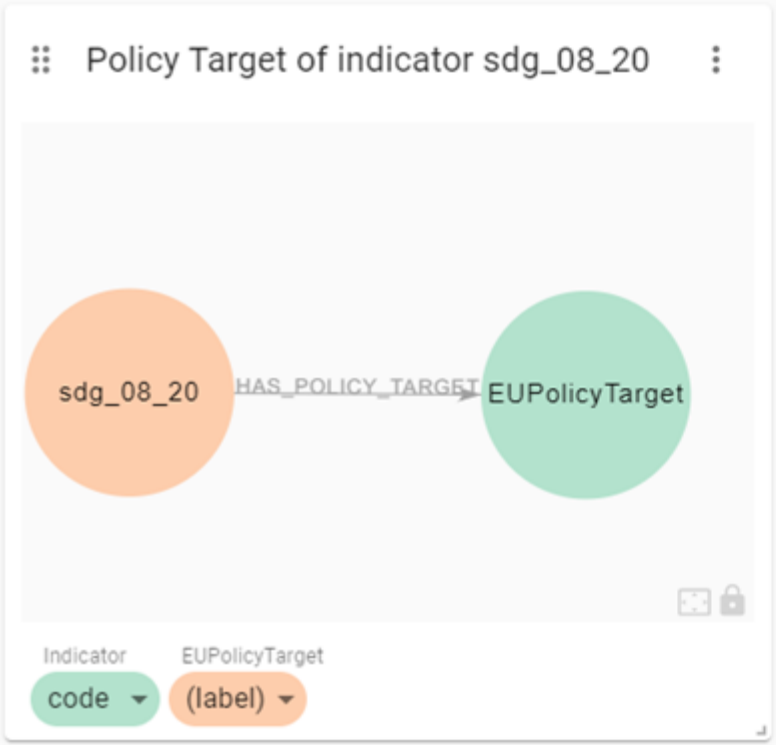
Indicative Visualizations (1/2)

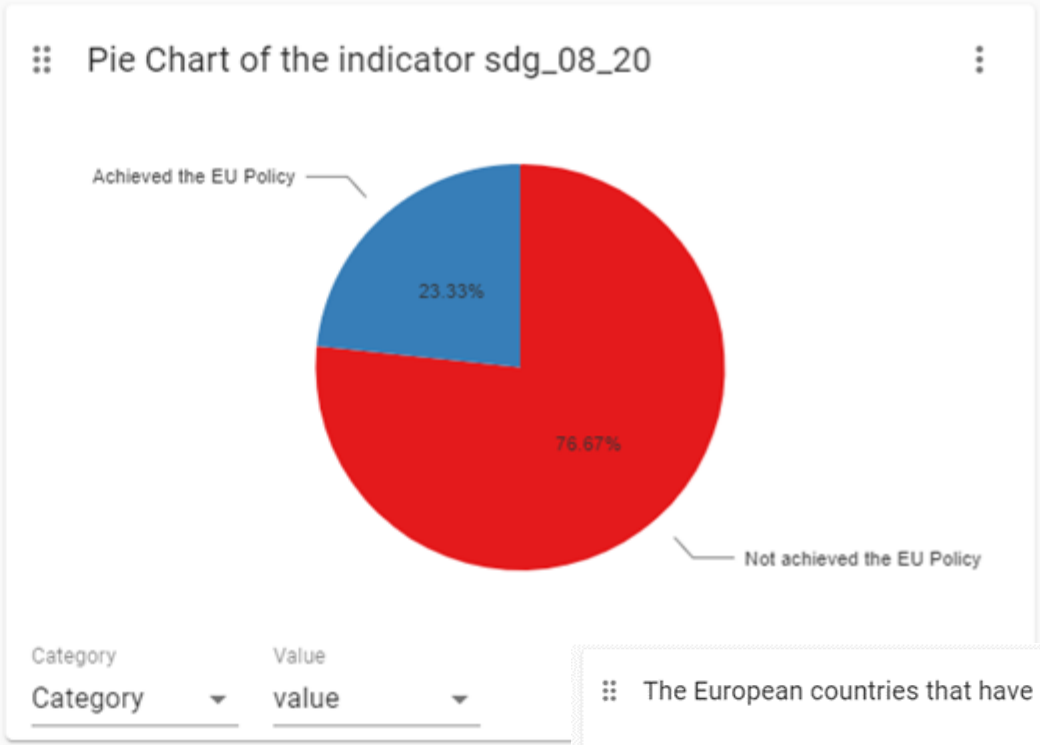


Stakeholders mapping in a case study in Athens



Indicative Visualizations (2/2)





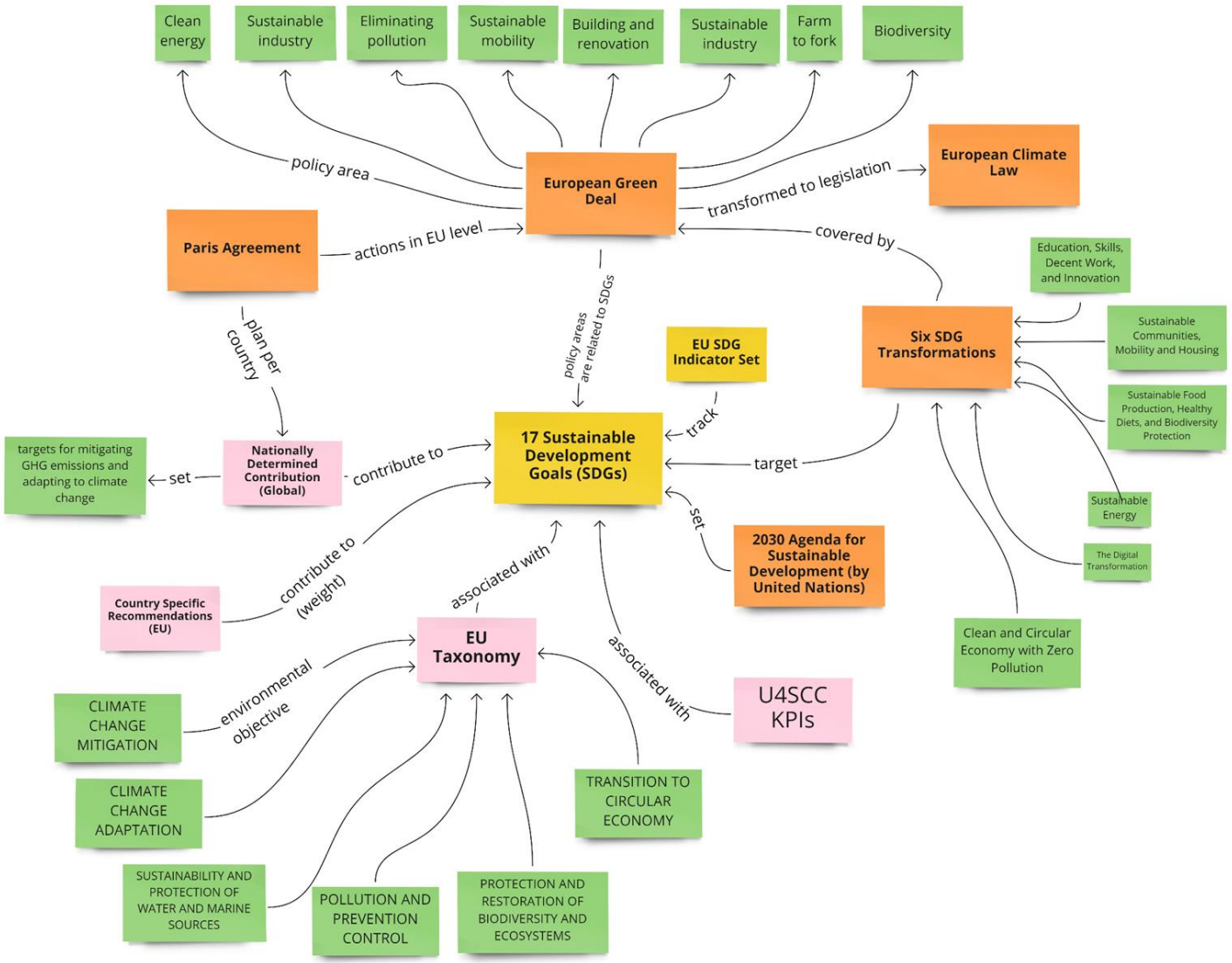
The European countries that have achieved or not the target

GeoArea	Status	Value
Netherlands	Achieved the EU Policy	5.5
Sweden	Achieved the EU Policy	6
Iceland	Achieved the EU Policy	7.3
Slovenia	Achieved the EU Policy	7.3
Norway	Achieved the EU Policy	7.4

1–5 of 30

Assessment for the achievement of a target for an EU SDG indicator

Mapping of policies documents to SDGs….not a straightforward task



Natural Language Processing for SDGs

Introduce Texts to the SustainGraph

Country **S**pecific **R**ecommendations:

- Documents prepared by the European Commission for each country analysing its economic situation and providing recommendations on measures it should adopt over the coming 12 months.

European **G**reen **D**eal Strategies

- Set of policy initiatives by the European Commission with the overarching aim of making the European Union (EU) climate neutral in 2050.

ARSINOE

NLP Techniques

Multi-label Classification using pretrained transformer-based models

Training dataset: **OSDG Community dataset** - https://github.com/osdg-ai/osdg-data

- Fine-tuning techniques on BERT, XLNet, GPT2 etc models
  - Train the entire architecture
  - Train some layers while freezing others
  - Freeze the entire architecture (*Transfer learning*)

Output: Probability scores with each SDG

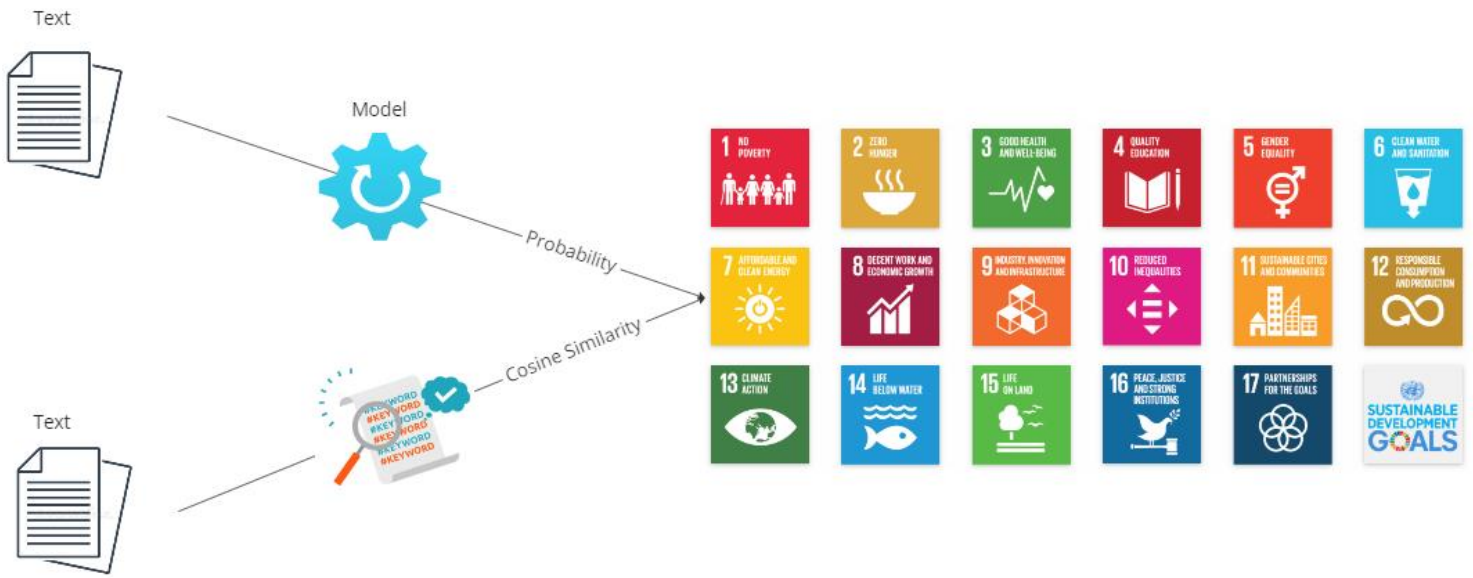
Cosine similarity between the top n keywords of texts with the keywords of the SDGs

- Find the candidate keywords/key phrases of the document (Tokenize and count the word occurrences)
- Convert the keywords/key phrases and the document to numerical data (Embeddings)
- Find the top n keywords of text based on the cosine similarity between the embeddings of text's keywords and document
- Compute the cosine similarity matrix between the top n keywords of text and the keywords of SDGs (https://sustainability.utoronto.ca/inventories/sustainable-development-goals-sdgs-keywords/ + https://ap-unsdsn.org/regional-initiatives/universities-sdgs/).

Output: Average Cosine Similarity Score with each SDG

ARSINOE

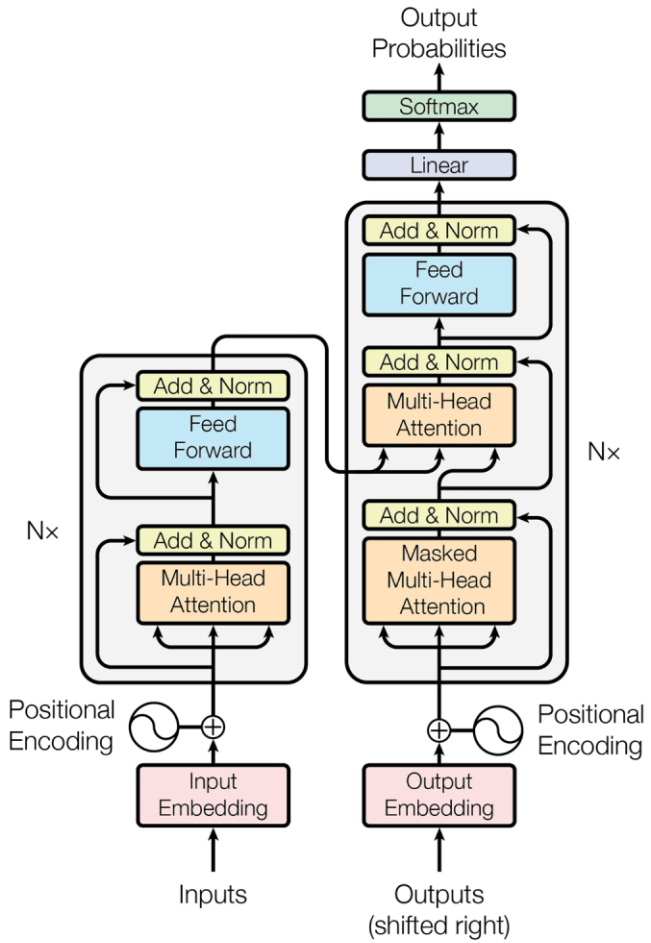
Combination of the two methods



$$r_{SDG} = \begin{cases} 0.7 * probability + 0.3 * avg\_cosine\_similarity & \text{,for SDGs 1-16} \\ 0.5 * avg\_cosine\_similarity & \text{,for SDG 17} \end{cases}$$

ARSINOE

Python Library SDGDetector

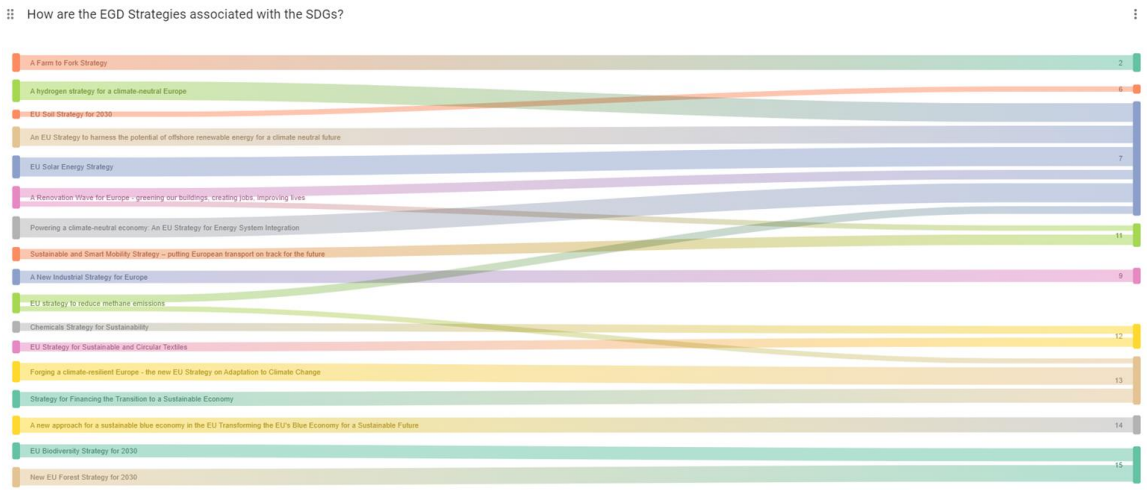


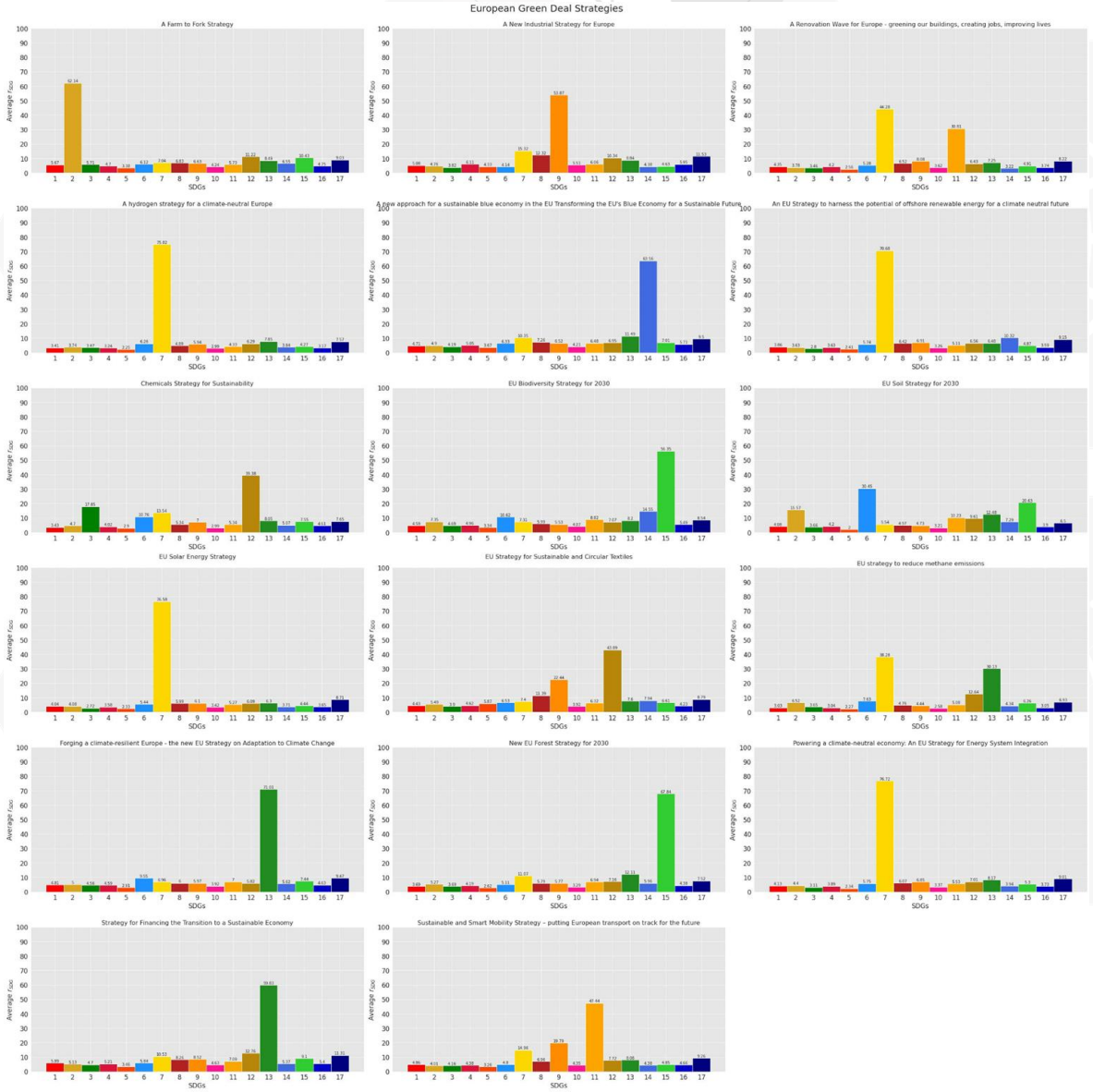
SDGDetector

Gitlab Repository **:** https://gitlab.com/netmode/sdg-detector/

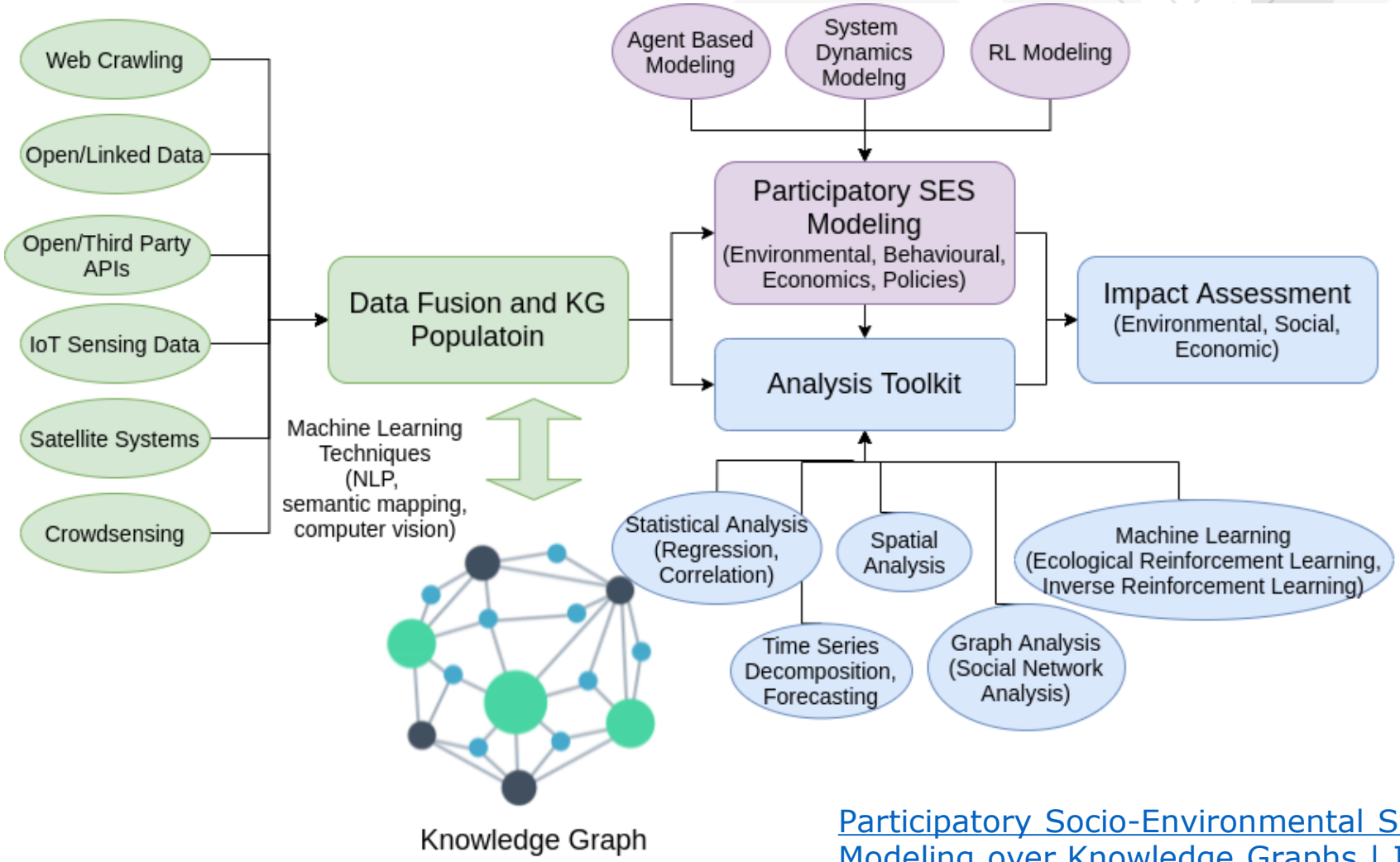
ARSINOE

SDGDetector Outcomes





Participatory socio-environmental systems modeling approach

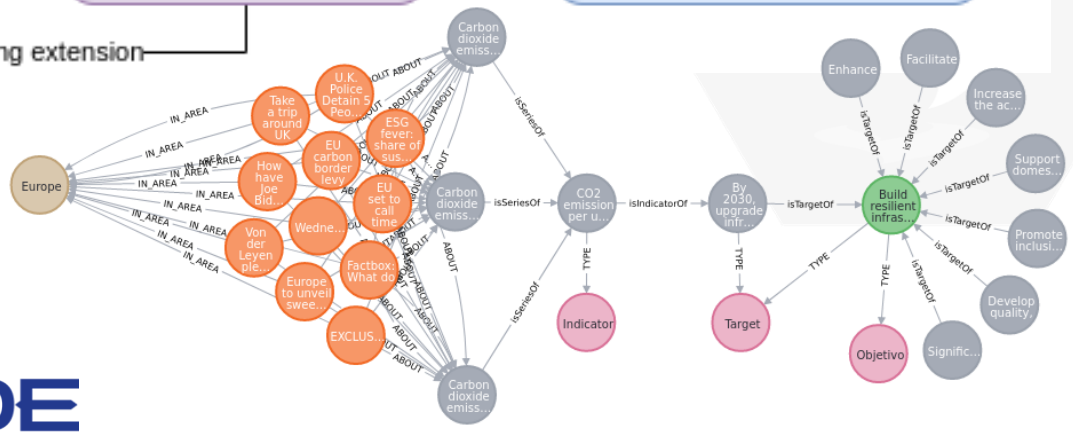


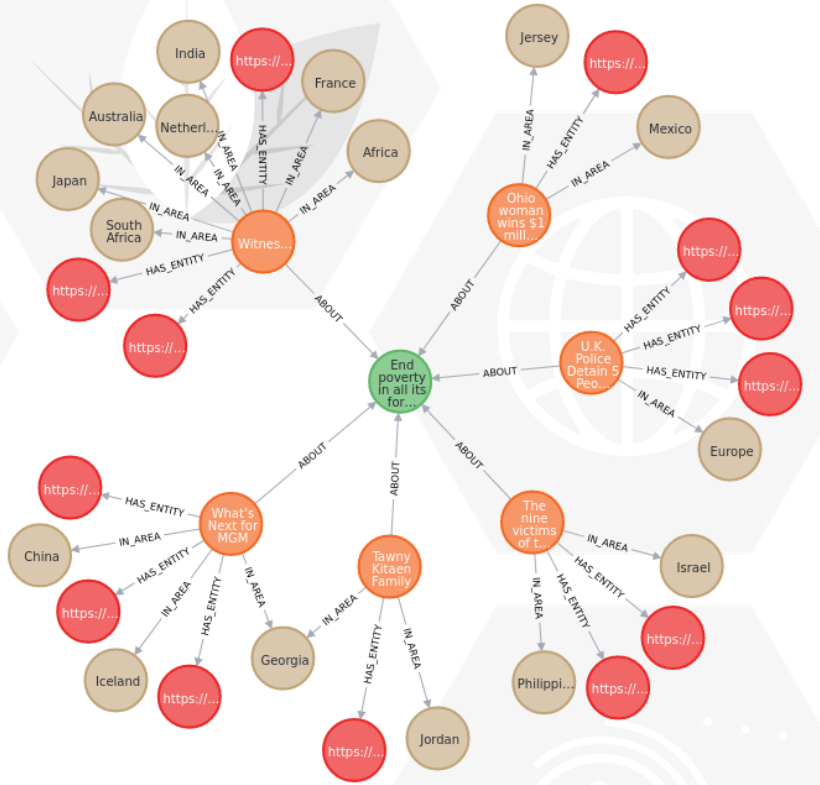
[Participatory Socio-Environmental Systems Modeling over Knowledge Graphs | IEEE Conference Publication]

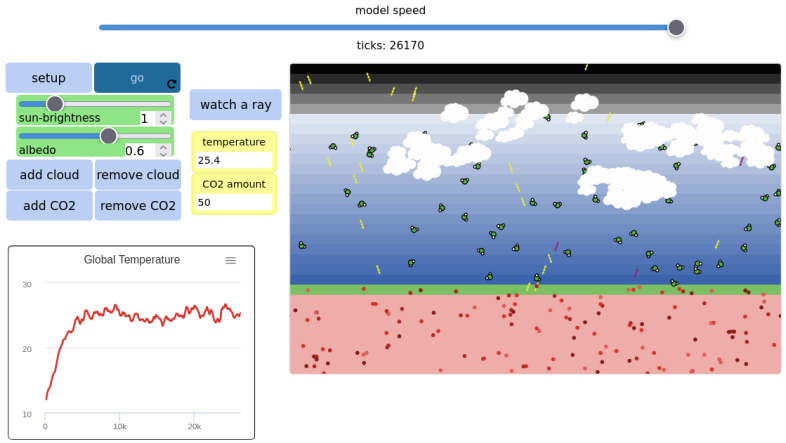
Indicative Usage Scenario

1. Import SDG Ontology Schema to KG
2. Add necessary prefixes and namespaces for representation of relationships in the KG
3. Import SDG Goals / Targets / Indicators / Series from United Nations SDG API and link them to SDG ontology
4. Import world wide news in the KG (web crawling)
5. Semantic enrichment of news articles via NLP entity extraction (ML Techniques)
6. Support queries related with SDG popularity across media in global, country and regional level
7. Adoption / Extension of Climate Change ABM
8. Feed Climate Change ABM with real SDG data
9. Creation of assessment reports per country

ABM modeling extension



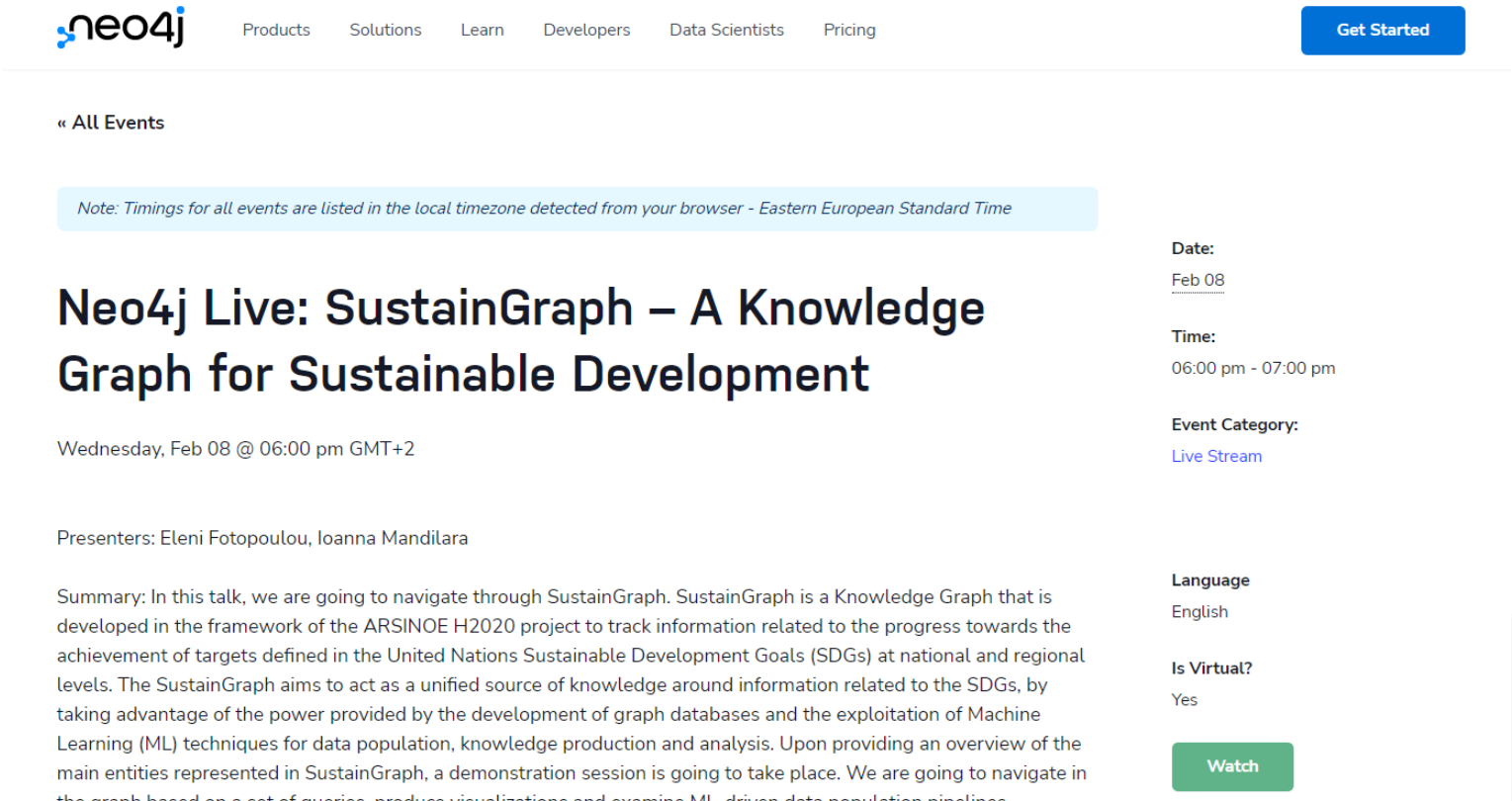




ARSINOE

For a live demo…join the Neo4j live session on Wednesday, 8th of February at 17:00 CET

https://neo4j.com/event/neo4j-live-sustaingraph/



THANK YOU

Anastasios Zafeiropoulos

tzafeir@cn.ntua.gr

ARSINOE

CLIMATE RESILIENT-REGIONS THROUGH SYSTEMIC SOLUTIONS AND INNOVATIONS

info@arsinoe-project.eu www.arsinoe-project.eu ARSINOE_EU

This project has received funding from the European Union's Horizon H2020 innovation action programme under grant agreement 101037424.