



# SciLake: improving open infrastructures for research assessment

CoARA-EOSC workshop, 10/4/2024 (online)

Thanasis Vergoulis, ATHENA RC



# At a glance

**Budget:** ~5M  
**Start date:** 1/1/23  
**Duration:** 36M

SKGs = Scientific Knowledge Graphs  
 infrastructures for research (meta)data

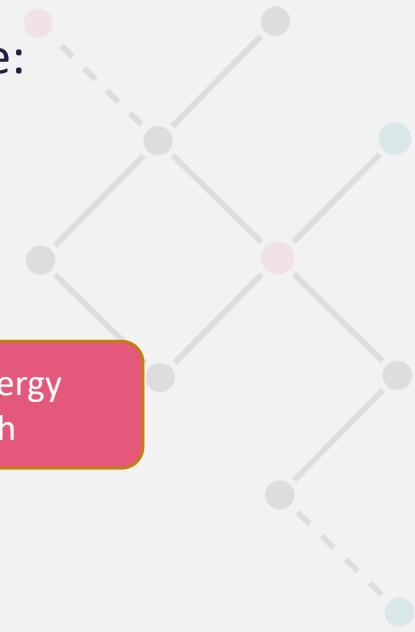
- **Scientific knowledge:** fragmented, heterogeneous (e.g., formats, languages)
  - Domain experts do not have the technical expertise in knowledge management
- SciLake introduces an **open & customizable Scientific Lake Service**
  - Hosts **open domain-specific & domain-agnostic SKGs\*** together with less structured domain knowledge.
  - Facilitates the **creation, interlinking & maintenance of community-created SKGs.**
  - Offers a **unified way to access & query** the contained assets.
- Demonstrate the concept with two **discipline-tailored, added-value services** to facilitate:
  - **navigating** the respective vast knowledge space (exploiting indicators of **scientific impact**)
  - improving **research reproducibility** in the respective domain
- Customise, test & demonstrate the developed services in **real-world pilots.**

**Pilot 1:** Neuroscience

**Pilot 2:** Cancer research

**Pilot 3:** Transportation research

**Pilot 4:** Energy research



## 9 countries / 13 organisations



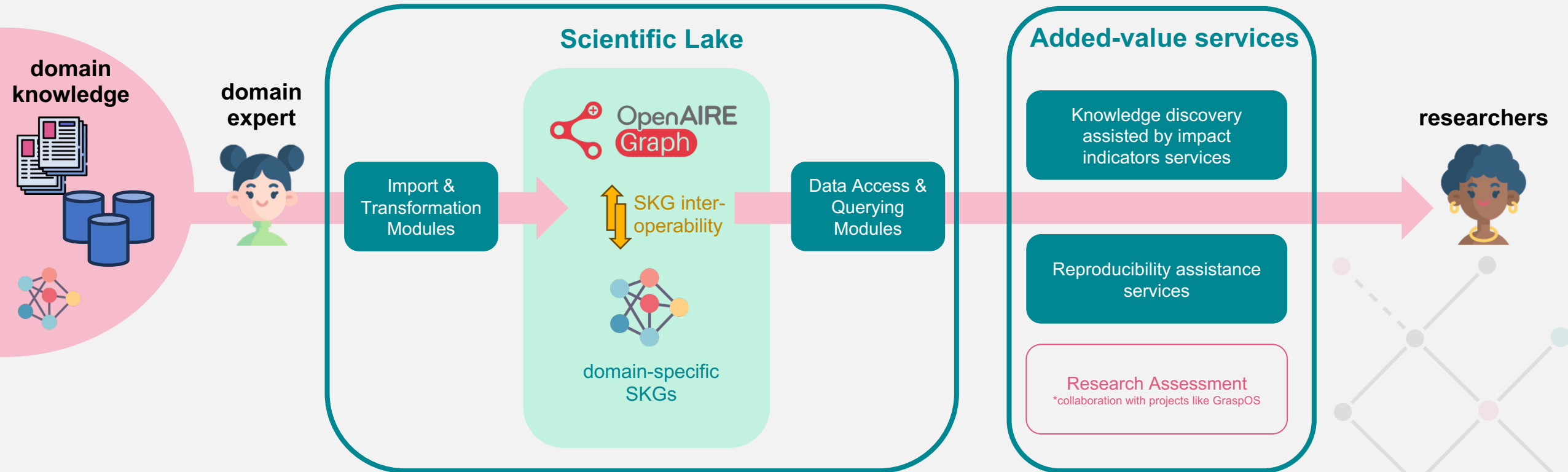
Map created from <https://www.mapchart.net>

**Coordinator**

EOSC & CoARA member

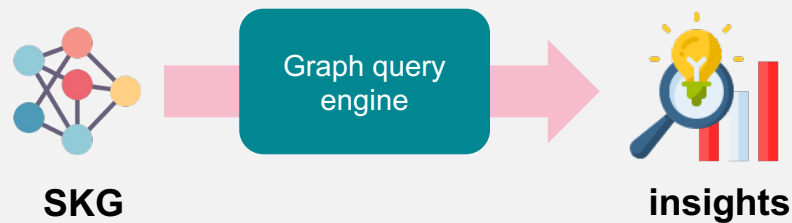
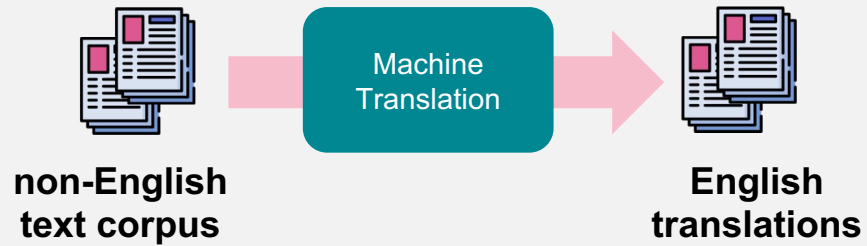
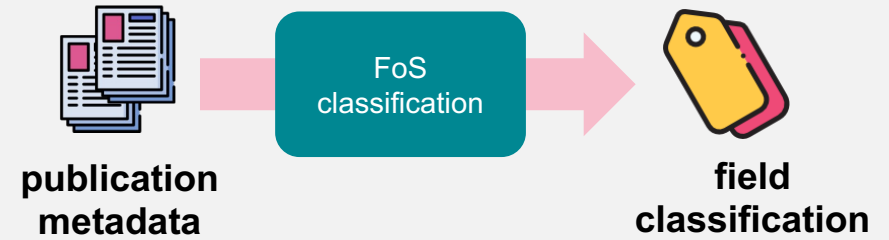
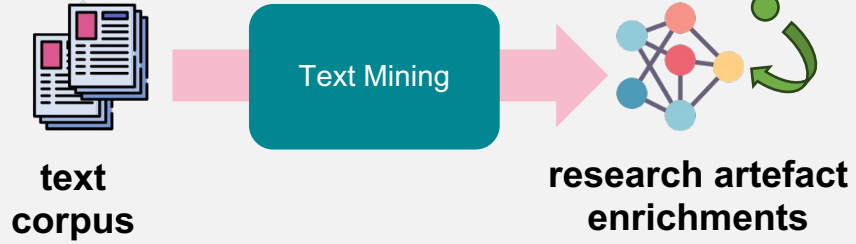


# The Concept



SKG "federation" ♦ multiple SciLake nodes deploying different modules

# Indicative Modules



# Contributions to Research Assessment

- SKG interoperability & accessibility ► easier to combine open SKGs & reveal insights
  - cooperation with RDA's SKG-IF group on model/API specifications to improve SKG interoperability
  - designing extensions ("profiles") of the basic SKG-IF model to support domain-specific entity & relationship types
  - enabling powerful graph querying of SKGs (via [AvantGraph](#) module)
- Modules to provide useful enrichments for research data
  - Fields of Science of research works (via SciNoBo FoS classifier)
  - Citation-based impact analysis/indicators of research works (via [BIP! services](#) & SIRIS)
  - Reproducibility analysis/indicators of research works (via SciNoBo & SIRIS)
  - In the life-time of the project, most of these contributions will be incorporated in the [OpenAIRE Graph](#)
- Acknowledgment of discipline-specific contributions
  - Examples: ML/AI models (from CS), protocols (from Life Sciences), curation contributions, etc.
  - These can be collected in domain-specific SKGs.

# SciLake

Contact us

Email: [vergoulis@athenarc.gr](mailto:vergoulis@athenarc.gr)

Twitter/X: [@vergoulis](https://twitter.com/vergoulis) ♦ [@SciLake\\_project](https://twitter.com/SciLake_project)

Thank you

Mastodon: [@vergoulis@scicomm.xyz](https://mastodon.social/@vergoulis@scicomm.xyz)

