# SciLake

Scientific Knowledge Graphs

**Title:** SciLake
**Project number:** 101058573
**Total budget:** € 4 809 450,00
**Duration:** 36 months
**Start Date:** 1 January 2023
**Project coordinator:** ATHENA Research Center (ARC)

## About SciLake project

### Scientific-Data-Lake-as-a-service

**Scientific Data Lake:** a data storage and analytics service to accommodate, manage, and expose heterogenous scholarly content through either domain-agnostic and domain **Scientific Knowledge Graphs (SKGs)** in conjunction with relevant unstructured data (e.g., **scientific texts**).

- Supports the creation and maintenance of domain-specific SKGs facilitating information extraction from scientific texts or consumption of transformed data from existing sources.
- Improves the interoperability of domain-specific or domain-agnostic SKGs.
- Enables easy and unified access to SKG contents, facilitating the implementation of added-value services.
- Exploits advanced automatic translation technologies to standardise multilingual content.

SciLake builds upon the **OpenAIRE Graph** to enrich it with domain-specific knowledge and pave the way for the creation of customisable added-value services for various research communities.

SciLake develops **customisable services** on top of the Scientific Lake to demonstrate its capabilities in **assisting knowledge discovery** and reproducibility taking into consideration specific domain requirements.

### Smart impact discovery service

Facilitates knowledge discovery in a particular scientific domain leveraging impact indicators:

- Estimates the impact of research objects (e.g., software, data) based on the impact of related publications.
- Utilizes classification schemes for scholarly publications to offer field-weighted impact indicators for research objects.
- Tracks research topics evolution and reveals relevant trends by quantifying the aggregated impact of each topic.

### Smart reproducibility assistant service

Helps researchers improve the reproducibility of their work:

- Identifies mentions of research objects (e.g., software, protocols, datasets) within publications offering insights about the reproducibility of the respective works.
- Reveals latent relationships among research objects.
- Informs researchers on the reproducibility of their works and offer suggestions on ways to improve it.

## Research community pilots

Four scientific communities participate in the evaluation and demonstration of SciLake services. For each, the developed services will be customised to accommodate differences in research procedures, practices, impact measures and types of research objects. The services will be validated and evaluated through real-world use cases.

**Neuroscience**

**Cancer**

**Transportation**

**Energy**

## Who benefits from SciLake?

| | |
|---|---|
| **Researchers** | SciLake supports researchers to create, manage, access, and process scholarly content with the purpose to enhance the visibility and the value of research outputs. |
| **Builders and operators of thematic or national SKGs** EOSC providers | SciLake improves the interlinking of SKGs and scholarly content through its tools and services, providing an easy, transparent, and effective management of scholarly content. |
| **Providers of scholarly communication graphs** | SciLake increases the offer of community-driven SKGs. |
| **Computer Science Communities** Big data, graph, AI, NLP, experts of standards | SciLake improves the quality of SKGs in terms of disambiguation and enrichment, impact and reproducibility assessment functions. |
| **Impact assessment experts** Research communities, SMEs | SciLake offers a multi-perspective analysis of impact metrics and indicators. It brings transparency in assessment exercises via the proliferation of open data and tools. |
| **Research administrators and policy makers** | SciLake enables identification of research objects, research trends, and effective assessment and tracking of the reproducibility level of research outputs. |
| **General public** | SciLake amplifies valuable and credible research in scientific communication. |

## Consortium

ATHENA Research & Innovation Information Technologies · OpenAIRE · icm · SIRIS A·C·A·D·E·M·I·C · Consiglio Nazionale delle Ricerche · ΙΕΣΛ ΕΠΙΣΕΥ · Karolinska Institutet · Opix Data Driven Insights · DFKI · University of Oslo · TU/e Eindhoven University of Technology · Hes·so · CERTH Centre for Research & Technology Hellas

Funded by the European Union

Supporting eosc

scilake.eu

@SciLake_project