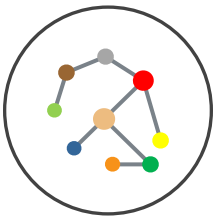
Connecting Science: Bridging Knowledge Borders Through Knowledge Graphs

What is SciLake?

As scientific publishing accelerates, one challenge grows steadily more complex: how to navigate and connect the expanding body of research outputs. Now, a European research initiative has taken the task to alleviate this problem. **SciLake** is a European-funded project that builds Scientific Knowledge Graphs (SKGs), which provide a graph database that facilitates the aggregation and querying of metadata from journal articles, datasets, methodologies and software, among other sources.



While SKGs are not new, SciLake brings two major contributions: domain-specific SKGs and the aggregation of multiple types of metadata from various repositories. To create domain-specific SKGs, SciLake has four thematic pilots in the fields of neuroscience, cancer, transportation and energy, where domain experts are building a highly customizable environment to generate knowledge graphs that are best suited to their specific needs.

The multilayer structure of these graphs provides more metadata than previous knowledge graphs, and more specifically, it enables users to identify questions in a meta-analysis style by highlighting research outputs based on their impact. Essentially, SciLake tackles the fragmentation and disorganization of scientific knowledge, which hinders the discovery of valuable insights, and facilitates informed decision-making in research.

From Neuroscience to Energy: Evaluating SciLake Across Domains

The SciLake project has developed four case studies, each focused on piloting domain-specific knowledge graphs. With **Athena Research and Innovation Center as the project's coordinator**, these pilots connect to several national and European initiatives adhering to specific scientific ecosystems.

Neuroscience

The **University of Oslo** and the **Research Infrastructure EBRAINS** have led the neuroscience pilot. Building on the OpenAIRE Graph, this pilot integrates metadata from EBRAINS Knowledge Graph and applies structured modelling approaches to tightly connect publications and datasets within a unified framework. Neuroscience data is particularly challenging to aggregate and interpret as it is methodologically diverse. This degree of data dispersity hinders its discoverability, a challenge SciLake can significantly improve.

Cancer

The cancer pilot has been built in collaboration with the **Karolinska Institutet** and the **Centre for Research and Technology Hellas,** and its primary focus is on structuring biomedical knowledge to support precision medicine. A domain-specific knowledge graph in this field allows the discovery of relationships between biomarkers, treatments, and clinical outcomes, which is particularly challenging to design treatments tailored to patients' individual needs and profiles. This approach enhances the accessibility and interpretability of complex cancer research, and supports more informed scientific and clinical decision-making.

Transportation

Led by the **Centre for Research and Technology Hellas** and the **National Technical University of Athens**, the transportation pilot explores how knowledge graphs can support innovation in mobility systems. This pilot aggregates fragmented research outputs from projects, datasets, and publications from two sub-domains: the Cooperative Connected Automated Mobility (CCAM) and Maritime Transport. The rapid evolution of terminology in these fields poses a challenge in experts' ability to stay informed and updated, and ensure regulatory compliance. The transportation implementation of SciLake enables stakeholders to identify technological trends, assess research impact, and uncover connections across the transport domain in a way that was not possible before. Ultimately, a more coherent and unified approach to data supports more sustainable transportation operations.

Energy

HES-SO Valais-Wallis is responsible for the energy pilot, which builds on ecosystems such as EnerMaps and connects multiple data sources to support energy system analysis, policy evaluation, and sustainability planning. Through the use of Scientific Knowledge Graphs and AI-driven services, it provides decision-makers with improved access to structured, cross-domain knowledge critical for managing the energy transition. In a time with growing climate challenges, the ability to implement effective changes in energy policies and keep the public informed is a crucial task that SciLake can help facilitate.

While the different pilots remain under development and user evaluation, the outcomes will reveal their feasibility and added value, especially in comparison to existing platforms and tools that are currently being used by researchers across the fields in which SciLake's graphs are being tested. One of the goals of the domain-specific SKGs usability testing is to improve data representation and its potential to be expanded to additional fields.

How does SciLake benefit researchers?

PhDs and postdocs

Navigating a complex and rapidly evolving field can be challenging. Early career researchers can benefit from using SciLake's SKGs for orientation and discovery in their research. These knowledge graphs facilitate the exploration of which datasets and methods are connected to influential papers and what techniques dominate in specific subfields, making them an impactful training tool for researchers' early career steps.

Principal Investigators and research leaders

SciLake's SKGs can potentially be used to track citations of established scientists' tools and datasets, and understand their position in the greater neuroscience landscape, identify relevant collaborators and facilitate grant strategy and creation of consortia. This information is significantly more complete than citation data alone, and it is becoming increasingly relevant for EU funding.

At a policy level, initiatives such as SciLake align closely with the principles of **Open Science** and **FAIR data** (Findable, Accessible, Interoperable, Reusable). As public investment in research increases, so does the need to make research outputs openly available and reusable. One of SciLake's long-term goals is to assess whether improved digital infrastructure can support researchers in making more informed decisions with regard to methods, collaborations and emerging directions.

Towards a more Open European Science Landscape

SciLake is only one example of a clear vision for European long-term strategic priorities: to offer solutions to critical challenges in research, innovation, and societal development. In a time with rapid geopolitical changes and where academic freedom and the protection of civil society seem under threat, European cooperation and commitment to open but safe data sharing is key. SciLake builds upon European Open Science Cloud (EOSC) functionalities and integrates its open-source services into the portfolio of EOSC core services.

Through collaborative European initiatives such as SciLake, the institutions involved contribute not only to scientific discoveries, but also to the evolving digital foundations that support them.

Contact:

Elisabet García González, PhD

Networking & Communications Advisor

University of Oslo | Neuroinformatics Norway | EBRAINS Norway

e.g.gonzalez@medisin.uio.no

UNIVERSITY OF OSLO

EBRAINS