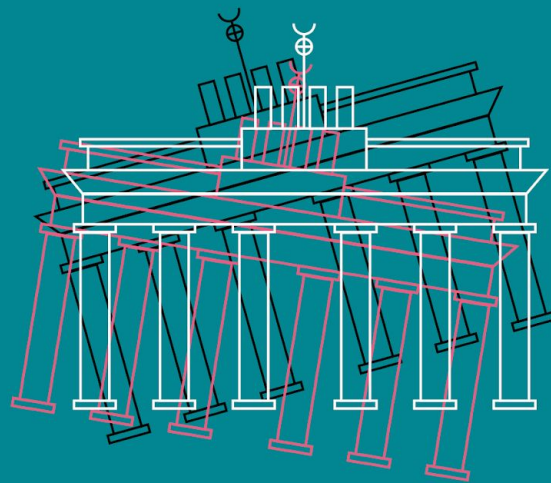




Unconference | EOSC collaborative frontiers to achieve interoperability and enhance scholarly data



EOSC Symposium 2024 / 21 - 23 October / Berlin, Germany

eosc.eu | [#eoscsymposium2024](https://twitter.com/eoscsymposium2024)



In cooperation with



EOSC Collaborative Frontiers to Achieve Interoperability and Enhance Scholarly Data

The Scientific Knowledge Graphs - Interoperability Framework (SKG-IF)

- ❑ Insights from research communities testing & implementing the framework
- ❑ Applications:
 - ❑ Research discovery
 - ❑ Data and metadata management
 - ❑ FAIRness, interconnectivity and machine actionability
 - ❑ Bibliometrics & research assessment
- ❑ Goal: Gather community feedback on SKG-IF development and foster collaborative efforts towards the EOSC-IF



Stefania Amodeo

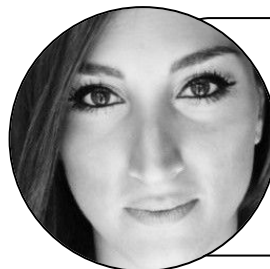
OpenAIRE
SciLake communication, engagement & exploitation
Session Chair



Thanasis Vergoulis

Athena Research Centre
SciLake & GraspOS project coordinator
RDA SKG-IF member

*The SKG Interoperability Framework:
progress and challenges in EOSC onboarding*



Elli Papadopoulou

Athena Research Centre
OSTrails deputy coordinator

*Steps towards FAIRness, interconnectivity and
machine actionability across all research phases*



Matt Buys

DataCite Executive Director
FAIRCORE4EOSC

*Insights from the network of repository
managers*



Andrea Mannocci

CNR-ISTI
GraspOS pilot on national research organisation
RDA SKG-IF co-chair

*Insights from research assessment and
bibliometric experts*



Giulia Malaguarnera

OpenAIRE
GraspOS WP4 coordinator

Discussion moderator: Next steps in EOSC

Join the discussion!

You can participate both [on-site](#) and [online](#)

1. Go to [slido.com](https://www.slido.com)
2. Use the code [#7111 054](#) or use the QR code
3. On-site participants: state in your question if you would like to ask your question live



WiFi network: EOSC

WiFi password: EOSCBerlin2024

The SKG Interoperability Framework: progress and challenges in EOSC onboarding

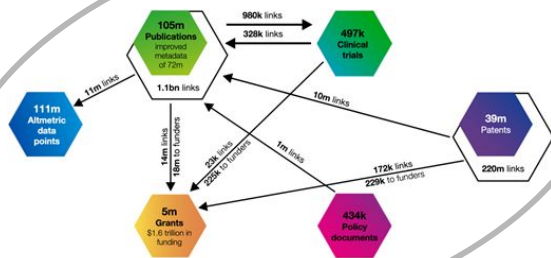
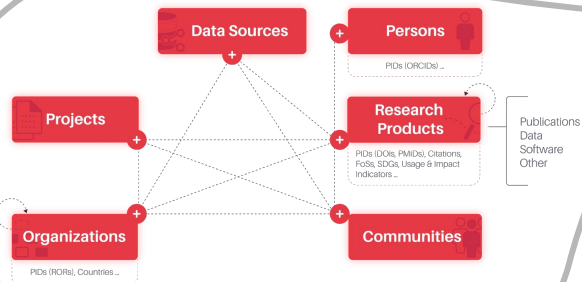


Thanasis Vergoulis

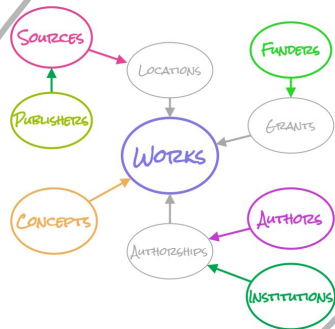
Athena Research Centre
SciLake & GraspOS project coordinator
RDA SKG-IF member

*The SKG Interoperability Framework:
progress and challenges in EOSC onboarding*

Background & Motivation



Status: September 2019



SKG-IF: At a glance

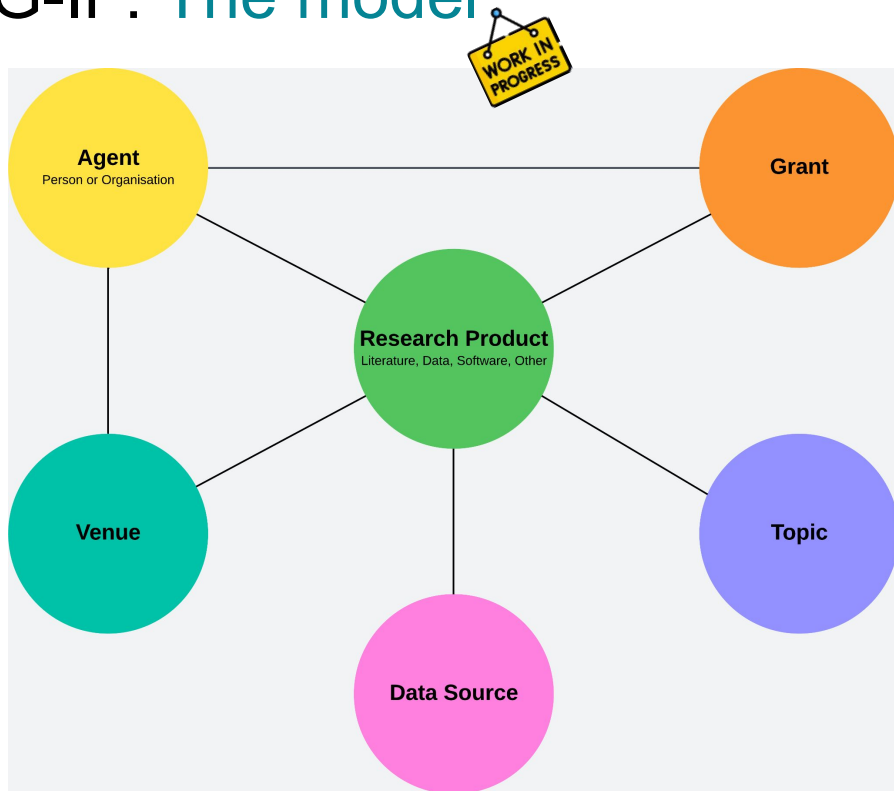
SKG-IF: Scientific Knowledge Graph Interoperability Framework

- Aim to define a metadata model to facilitate
 - Interoperability between SKGs
 - SKG content consumption by value-added services developers
- Inspired by the models of existing, well-known SKGs
- Intention to be part of the EOSC Interoperability Framework

More information:

- <https://skg-if.github.io>
- <https://skg-if.readthedocs.io/en/latest/> (old website, being discontinued with redirect)

SKG-IF: The model



*Documentation sample

interoperability-framework

Agent

The **Agent** entity represents an individual (e.g., a person, an organisation, or another kind of entity being able to act) who is involved in the creation, publication, dissemination, etc. of a **Research product**. An **Agent** can be an author, a reviewer, an editor, a publisher, a researcher, or any other stakeholder involved in the scholarly communication process.

This page describes the metadata fields for an **Agent** and all its subtypes (i.e., persons and organisations).

Properties

local_identifier

String (mandatory): Unique code identifying the **Agent** in the SKG (if any, otherwise "stateless identifier").

Suggestion: use a URL as a string to make this resource dereferenceable on the Web.

```
"local_identifier": "skg-if-wg"
```

identifiers

List (recommended): A list of objects representing external identifiers for the entity. Each object is structured as follows:

- *scheme* *String* (mandatory): The scheme for the external identifier (e.g., orcid, viaf, etc.).
- *value* *String* (mandatory): The external identifier.

Note: the current version of SKG-IF includes the following types of identifiers (to be specified as strings in the field "scheme"): orcid, viaf, ...

```
"identifiers": [  
  {  
    "scheme": "orcid",  
    "value": "0000-0002-5193-7851"  
  }  
]
```


SKG-IF: History

- Incubated in **RDA IG Open Science Graphs for FAIR data**
 - <https://rd-alliance.org/groups/open-science-graphs-fair-data-ig>: Helena Cousijn (DataCite), Amir Aryani (ResearchGraph.org), Paolo Manghi (OpenAIRE), Andrea Mannocci (CNR-ISTI/OpenAIRE), Anita de Waard (Elsevier)
 - Self-organised in Task Forces since June 2022
- Transitioned to **RDA WG Scientific Knowledge Graphs - Interoperability Framework**
 - <https://www.rd-alliance.org/groups/scientific-knowledge-graphs-interoperability-framework-skg-if-wg>
 - Officially endorsed in June 2023, wrap-up in December 2024 (18 months)
 - Chairs
 - **Andrea Mannocci** (CNR-ISTI / OpenAIRE)
 - **Silvio Peroni** (OpenCitations)
 - **Jason Portenoy** (OpenAlex)
 - **Sahar Vahdati** (InfAI)

SKG-IF: Roadmap

SKG Core Information Model

Review the models of a selection of SKGs and checked for common research entities being modelled and the respective properties. Trying to keep a **balance between pragmatism and expressivity** of the model.

Ongoing maintenance and extension

SKG Data Exchange Commons

Review **metadata formats** that could be suitable for data exchange

Transitioning to JSON-LD

SKG Access Protocol Commons

Review **technical protocols** and means to achieve information exchange. From simple data dumps to API and querying.

Defined an API resolver

Full-fledged API yet to be discussed

SKG Extensions

Setup a **mechanism to extend the basic model** to cover special needs of communities and use cases.

Defined rules of engagement

SKG-IF: Extensions

The **core model** should be capable of modeling the most **widely used research-related entities**.

The focus should be on **domain-agnostic entities, relationships, and properties**.

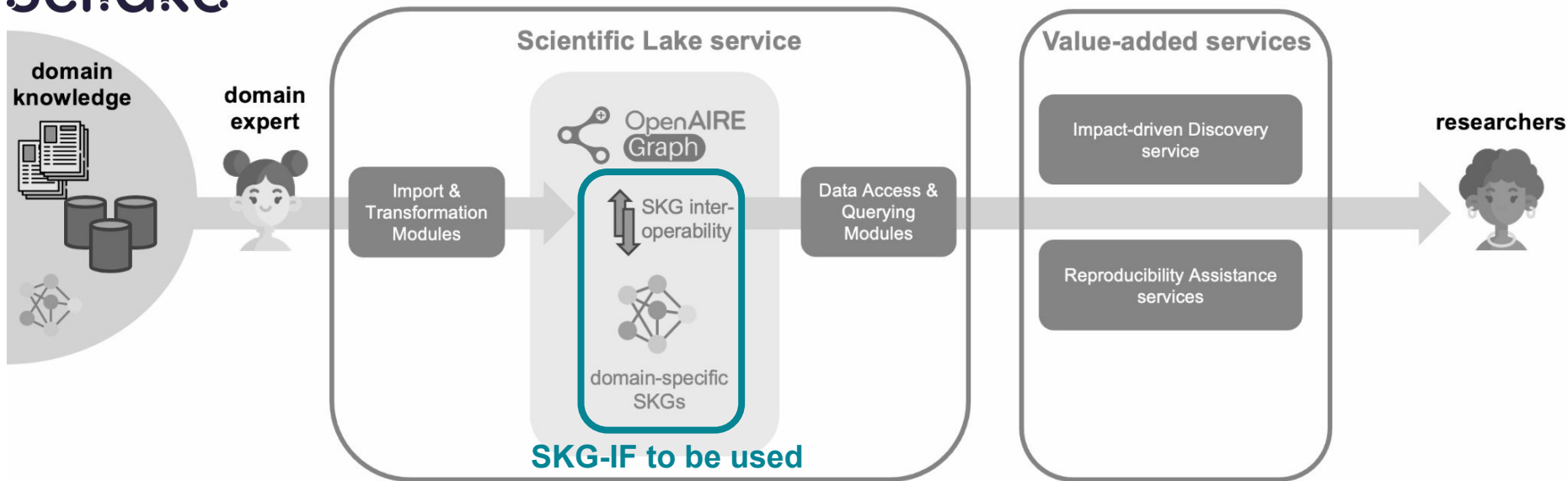
**Impossible/unsustainable to keep up-to-date domain-specific knowledge.*

Need to support an **extension mechanism** to cover relevant needs.

Official extensions to be **listed on the SKG-IF website** along with documentation.

Motivation example for SKG-IF extensions

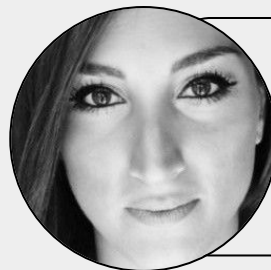
SciLake



SKG "federation" ♦ multiple SciLake nodes deploying different modules

Icons created by ultimatearm, Becris and Freepik from www.flaticon.com.

Steps towards
FAIRness,
interconnectivity and
machine actionability
across all research
phases



Elli Papadopoulou

Athena Research Centre
OSTrails deputy coordinator

*Steps towards FAIRness, interconnectivity and
machine actionability across all research phases*

Outline

- Overview of the OSTRails project
- Draft IFs for SKGs, DMPs and FAIR & proposed extensions
- Reference Implementations by pilots
- Next Steps: towards “Commons”



Definition

In OSTRails: an SKG is defined as any database/repository with information pertaining to research products, process and actors & agents which is able to present a graph type view on this information via a suitable API.

Today's limitations

DMPs

- No evidence of better RDM or FAIRer results
- DMPs not published or FAIRified
- Uncertainty of the process
- DDPs pending adoption



maDMPs

Describes DMP entities and their links with outside sources

SKGs

- Mostly for bibliometrics & discovery
- Isolated
- Quality issues
- Depth of knowledge in communities



maDOs

Describes elements of digital objects and links with other resources (qualified references)

FAIR

- Policy vs Implementation
- Inconsistent results between tools
- No FAIR assessment schema
- FAIR vs Data Quality

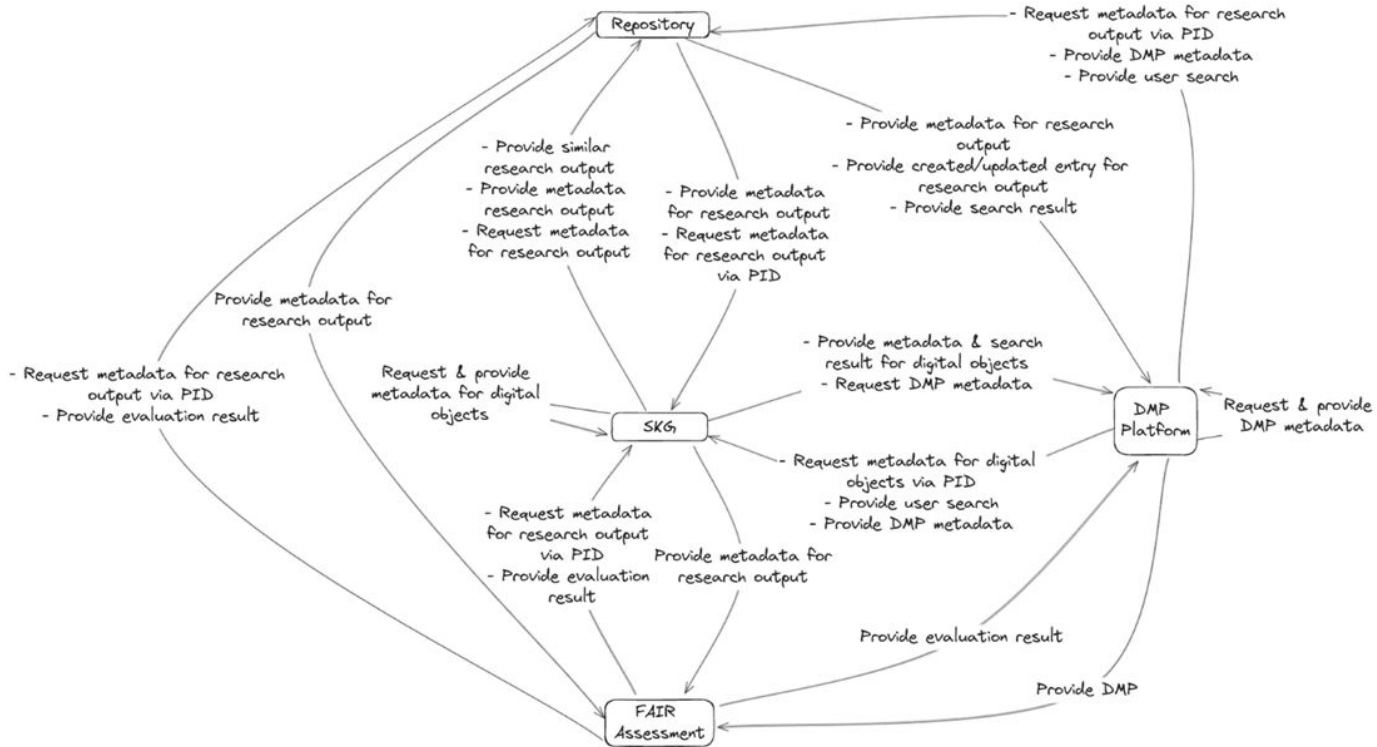


maFAIRTests

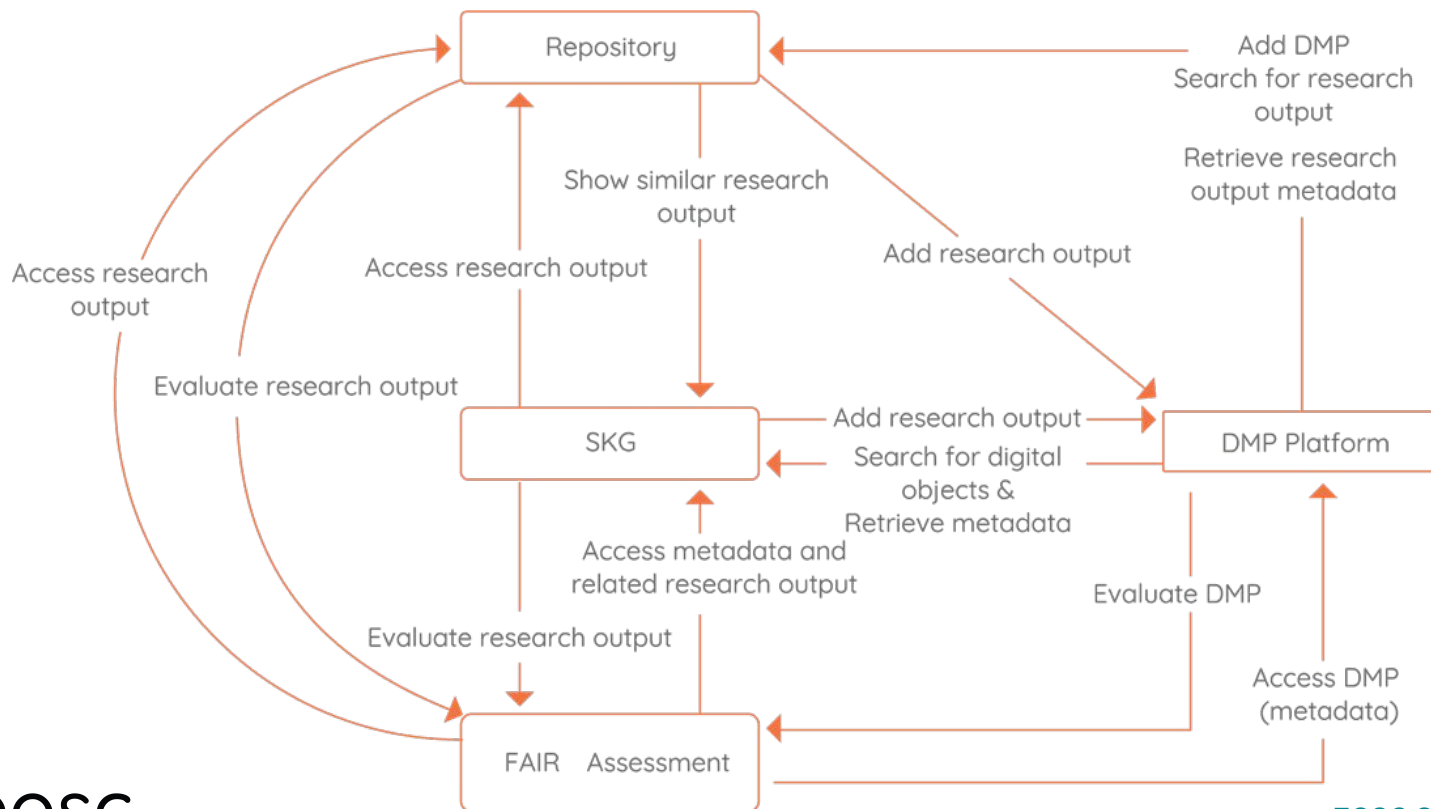
Describes community FAIR rules for each element in DOs to assess its overall FAIRness

Pathways

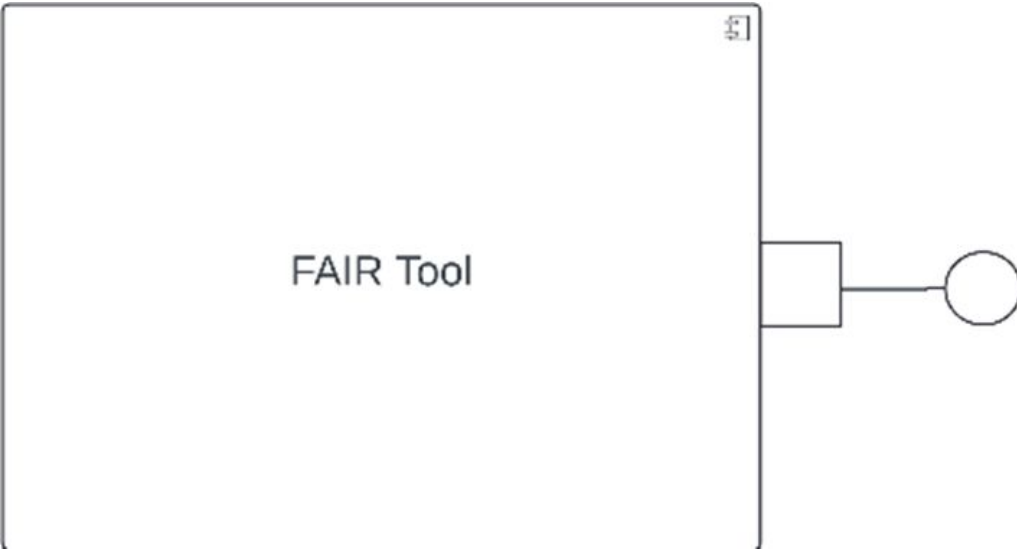
Tools Actions



Plan-Track-Assess Framework



Draft Interoperability Frameworks



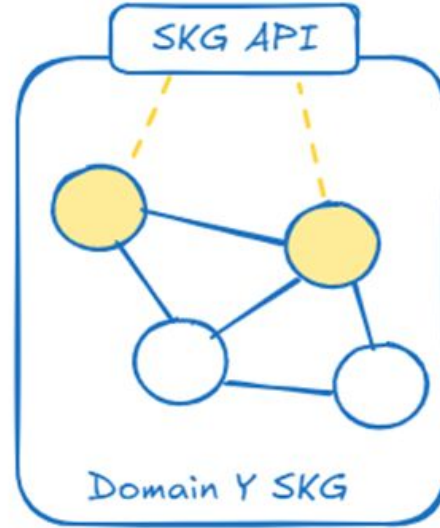
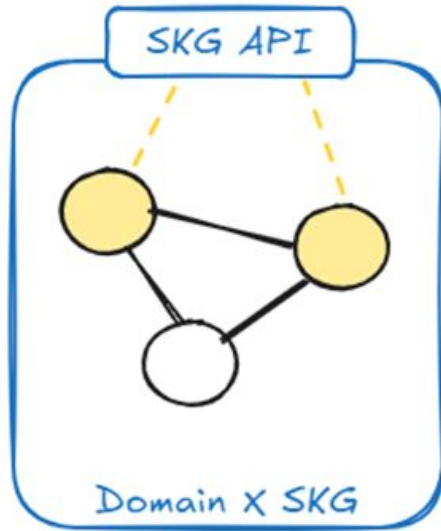
Provided Interfaces:
Expose Test Execution
Expose Test Metadata
Expose Metrics involved in FAIR Assessment

Pilots - Reference Implementations

- Expose data according to SKG-IF
- Enhance SKG-IF with new entities and semantics, incl. qualified references
- Quality of data: resource_types & annotations
- Rich metadata



OTrails Commons





Walking the pathways together



Insights from the network of repositories



Matt Buys

DataCite Executive Director
FAIRCORE4EOSC

*Insights from the network of repository
managers*

Context & Perspective

DataCite

DataCite is a global community of research organisations that shares a common interest: to ensure that research outputs and resources are openly available and connected so that their reuse can advance knowledge across and between disciplines, now and in the future.

As a global non-profit membership organization, we work with 3500+ repositories in the world to provide DOIs for research outputs and resources.

FAIRCORE4EOSC

Developing EOSC-Core components to enable a FAIR EOSC ecosystem

The FAIRCORE4EOSC project focuses on the development and realisation of core components for the European Open Science Cloud (EOSC). Supporting a FAIR EOSC and addressing gaps identified in the Strategic Research and Innovation Agenda (SRIA). Leveraging existing technologies and services, the project will develop nine new EOSC-Core components aimed to improve the discoverability and interoperability of an increased amount of research outputs.

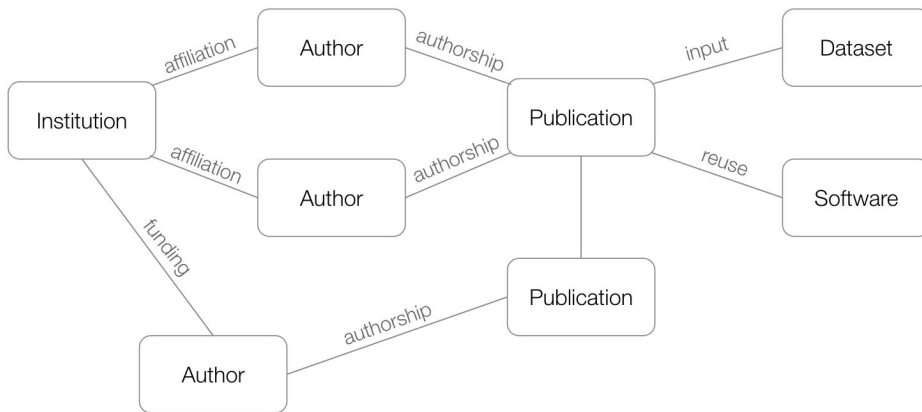
FAIRCORE4EOSC has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101057264

PID Graph Background

PID Graph – Concept. The connected searchable Graph of publications, datasets, other research outputs, people and organizations.

PID Graph – User Stories. At the beginning of the FREYA project we collected more than 40 user stories (see <https://pidforum.org/c/pid-best-practices/8>) that were difficult to address with currently available services.

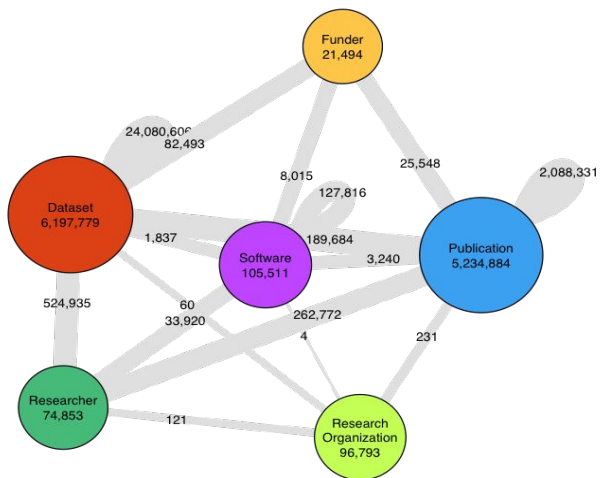
PID Graph – Architecture. We discussed and prototyped PID Graph architecture to address user stories, taking into account existing PID infrastructure, in particular from unfunded partners, and sustainability beyond FREYA project duration.



PID Graph Scaling

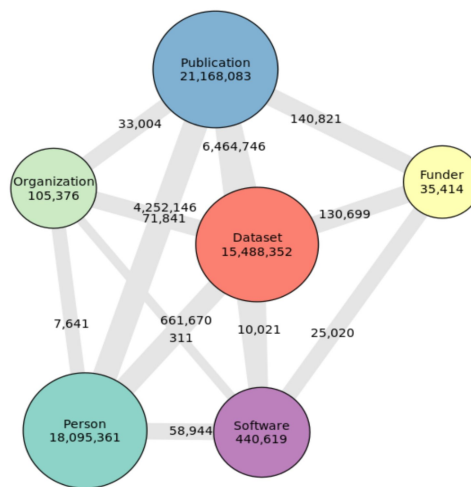
PID Graph

Nodes and connections (as of September 2019)



PID Graph

Nodes and connections (as of October 2023)



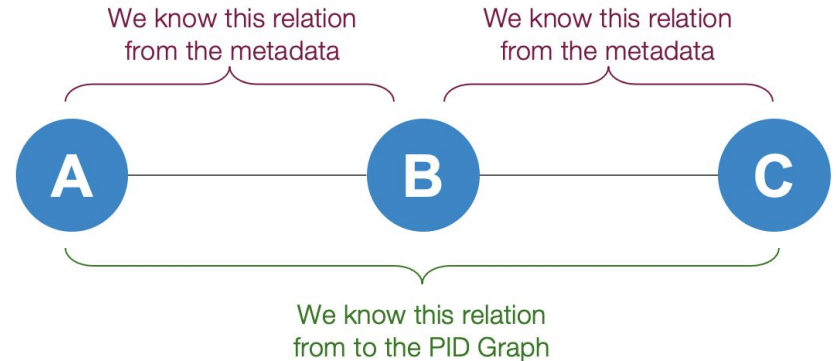
Insights: Repositories

Repositories' persistent identifiers (PIDs) and metadata feed into the PID Graph, helping to connect research outputs and resources.

FAIRCore4EOSC is making it easier to harvest metadata and understand relationships between hundreds of millions of entities.

Unique PIDs for researchers and outputs are key to linking pieces of the research landscape.

By linking PIDs through metadata relations, we can now discover connections at least two “hops” away. Bringing rigor to the scholarly records.



SKG-IF: Our Plans

The Scientific Knowledge Graphs Interoperability Framework (SKG-IF) is maturing, with a focus on adopting a standardized approach using JSON-LD for representing metadata.

We are committed to exploring how to implement and contribute to this initiative, ensuring alignment with our goals of improving research discovery and accessibility.

We advocate for interoperability that goes beyond standard frameworks, aiming for solutions that enhance metadata exchange across the full research lifecycle.

As part of our commitment to open metadata, we plan to offer an annual data dump in SKG-IF format, enhancing access to interconnected research information.

- Additionally, there is potential to extend this offering through a REST API, allowing for content negotiation and seamless data integration across platforms.

Moving forward, we will focus on planning the effective implementation of an API to support broader adoption and use of the SKG-IF format.

PID Graph: Impact for Repositories

Enhanced Reuse:

Regular data dumps related to community profiles, available through the PID Graph, support repositories in contributing to the EOSC community and beyond.

Streamlined Ingestion:

The API and data dumps, featuring DOIs and metadata from DataCite, ensure seamless integration into EOSC and related services.

Data Usage and Stats:

Integrated statistics on DataCite DOI usage help repositories track and analyze the reach and impact of their content.

Contextual Metadata:

By linking PID entities, the PID Graph enriches repositories' metadata, uncovering connections between research outputs and resources, affiliations, and contributors.

Aggregated PID Links:

Supporting PID link aggregation helps repositories highlight the interconnectedness of research, fostering deeper insights and visibility.

Insights from research assessment and bibliometric experts



Andrea Mannocci

CNR-ISTI
GraspOS pilot on national research organisation
RDA SKG-IF co-chair

*Insights from research assessment and
bibliometric experts*

Context

GraspOS - Next Generation Research Assessment to Promote Open Science (<https://graspos.eu>)

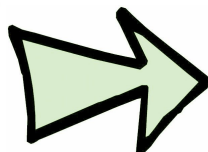
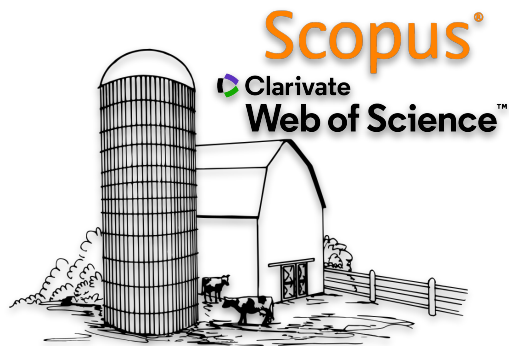


OI4RRA - COARA WG Towards Open Infrastructures for Responsible Research Assessment (<https://coara.eu/coalition/working-groups/wg-towards-open-infrastructures-for-responsible-research-assessment-oi4rra>)



Both contexts focus on the need for and rise of **Open Infrastructures for Research Assessment**

Open Infrastructures for Research Assessment



Intended usage of SKG-IF

- Foreseen as the identified **backbone** of such infrastructures
- *"Lingua franca"* that glues together different building blocks and enable exchange of information across federated data sources and services
- Some **SKGs already committed** (OpenAIRE, DataCite, OpenCitations), others contributed and are checking our work (Crossref, OpenAlex)



Current limitations of SKG-IF towards RA

In the current iteration of the SKG-IF, a **few aspects** that are crucial for research assessment **are missing**

- The fields describing entities in the SKG-IF Core model are currently **limited to bibliographic metadata**
- No coverage of potentially relevant **indicators** to quantify impact and usage
- No inclusion of **narratives** providing context to contributions to science

Current work extending SKG-IF

Within **GraspOS**, we formed a WG discussing on the **SKG-IF extension**

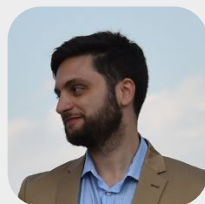
- **Specification of indicators as a "decorator" of the SKG-IF Core entities**
 - Build on previous experience from data sources sharing this kind of information
 - Identify and prioritise **entities to target** (products, actors, organisations, grants, etc...)
 - Define an **adequate data model for indicators** enabling explainability (e.g., how a given indicator has been calculated, using which data, under which circumstances, validity, etc...)
- **Definition of the data model for CV narratives at multiple levels**
 - Single research products
 - Streams of research products
- **Customisation of core API specifications**
 - Core APIs are bare-bones (i.e., a resolver with optional HTTP parameters)
 - Play with HTTP parameters and see if this can cover all the requirements
 - Optionally add new REST methods to cover what is needed



Moderated discussion



**Giulia
Malaguarnera**
(Chair),
OpenAIRE



**Thanasis
Vergoulis,**
Athena Research
Centre



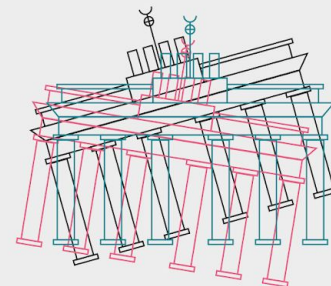
Elli Papadopoulou,
Athena Research Centre



Matt Buys,
DataCite



**Andrea
Mannocci,**
CNR-ISTI



Join the discussion!

You can participate both [on-site](#) and [online](#)

1. Go to [slido.com](https://www.slido.com)
2. Use the code [#7111 054](#) or use the QR code
3. On-site participants: state in your question if you would like to ask your question live



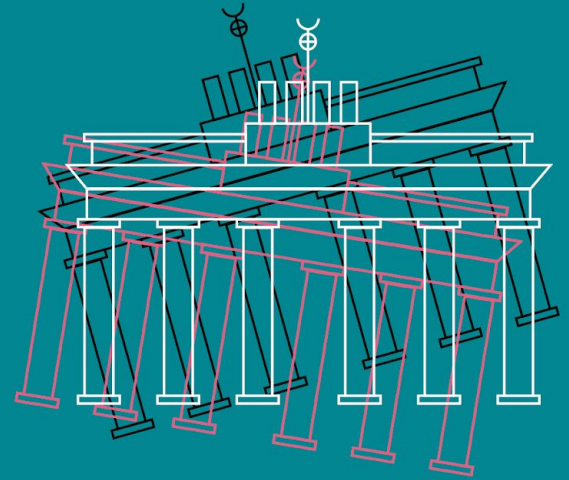
WiFi network: EOSC

WiFi password: EOSCBerlin2024



Networking lunch 12:30 – 14:00

Next session at 14:00
Legal issues along the research data cycle



EOSC Symposium 2024 / 21 - 23 October / Berlin, Germany