Scientific Knowledge Graphs Interoperability Framework (SKG-IF): Enabling Data Exchange for SKGs

A Report from the RDA SKG-IF Working Group

Andrea Mannocci[1], Silvio Peroni[2], SKG-IF WG[3]
[1] CNR-ISTI, Pisa, Italy
[2] University of Bologna, Bologna, Italy
[3] Research Data Alliance (RDA)

Abstract

The SKG-IF aims to facilitate **seamless exchange of information across diverse Scientific Knowledge Graphs (SKGs)**, thereby reducing duplication of effort, enriching individual graphs, and establishing SKGs as a central pillar of Open Science scholarly communication. This report outlines the work undertaken in this direction by the Research Data Alliance (RDA) Working Group on the Scientific Knowledge Graphs - Interoperability Framework (SKG-IF). It addresses the increasing need for **interoperability among SKGs** to support Open Science and FAIRness in research, and foster the transition to Open Research Information (ORI), in compliance with the recent Barcelona Declaration. This report details the context, motivation, and goals behind the development of the SKG-IF, presents its core components including the **data model (SKG-IF Data Model)**, **API specifications**, **and extension guidelines**, and provides pointers to the comprehensive online documentation, which ultimately remains the most up-to-date and detailed source of information about the framework.

1. Introduction

The Open Science movement is driving a paradigm shift towards more open and accessible research outputs, including data and software, alongside traditional publications. This has led to the proliferation of specialised **Scientific Knowledge Graphs (SKGs)** designed to meet various stakeholder needs, such as discovery, reuse, and impact assessment. These SKGs offer rich metadata representations about research entities, such as people, research products beyond literature items, institutions, and the relationships between them.

However, SKGs are typically devised and evolve with a specific use case in mind, and this translates to pervasive data heterogeneity and localised data models. While individually valuable, such a fragmentation hinders the full potential of modern SKGs, as incompatible systems render cross-domain synthesis and analysis prohibitively complex.

Recognising this challenge, the RDA Interest Group (IG) on Open Science Graphs for FAIR Data[1] initiated work towards an Interoperability Framework for SKGs. Building on this foundational effort, the RDA Working Group (WG) on Scientific Knowledge Graphs - Interoperability Framework[2], endorsed in June 2023, has focused on defining a framework to enable the **seamless exchange of information across these diverse graphs**. This report presents the key outcomes of this WG's efforts.

2. Motivation and Goals

The primary motivation for the SKG-IF stems from the need for **SKGs to interoperate** to support Open Science and improve FAIRness in research. As highlighted in Aryani et al. (2020), interoperability can **reduce duplication of effort and capitalise on synergies and complementarity across diverse SKGs**. It also **enables information to circulate, enriching individual SKGs and providing redundancy** to safeguard information availability and persistence. Ultimately, the goal is to **elevate SKGs as the backbone of Open Science scholarly communication,** and foster the transition recently promoted by the Barcelona Declaration[3] towards Open Research Information (ORI).

The SKG-IF aims to mediate the diverse data models and technologies currently employed by existing SKGs. By establishing a **community-driven framework**, the WG sought to:

- Define a common understanding of **six core research entities, their relationships and attributes**;
- Provide a **core data model (SKG-IF Data Model)** that can represent key information in a **standardised manner**;
- Outline **API specifications** to facilitate **programmatic access and exchange** of SKG-IF compliant data;
- Enable the specification of **extension of the core data model and the API** to accommodate specific community needs without compromising interoperability.

The WG recognised that a single "**lingua franca**" might be too restrictive given the complexity and diversity of SKGs, and thus focused on enabling exchange at both the **information model and technological levels**.

[1] https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig
[2] https://www.rd-alliance.org/groups/scientific-knowledge-graphs-interoperability-framework-skg-if-wg
[3] https://barcelona-declaration.org

3. Core Components of the SKG-IF

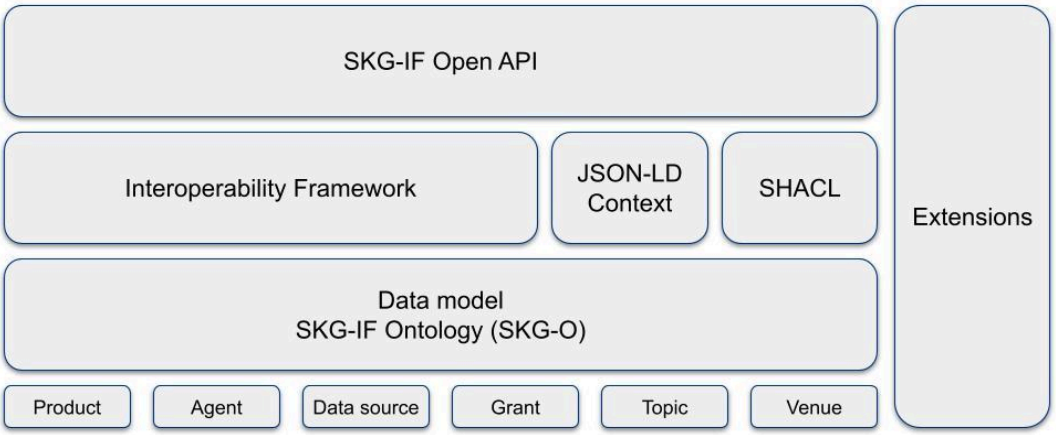

Figure 1 - SKG-IF architecture.

3.1. Data Model

The SKG-IF **Data Model** provides a formal and ontologically sound definition of the data represented by the SKG-IF Interoperability Framework. The SKG-IF facilitates the exchange of data about **six core entities, their relationships and attributes**:

- **Research product:** Encompasses research literature, research data, research software, or other uniquely identified digital assets;
- **Agent:** Represents an individual (e.g., a person) or organisation involved in the creation, publication, dissemination, etc., of a Research product;
- **Grant:** Describes funding awarded to an Agent by a funding body;
- **Venue:** Models a publishing "gateway" used by an Agent to make their Research products available;
- **Topic:** Describes the scientific disciplines, subjects, and keywords potentially relevant for a Research product;
- **Data source:** A service or platform where a Research product (its metadata and files) is stored, preserved, and made discoverable and accessible.

The data model employs the "language" of the Semantic Web and reuses existing ontological models, including the Semantic Publishing and Referencing (SPAR) Ontologies[4] (Peroni & Shotton, 2018), the Provenance Ontology (PROV-O) (Lebo et al., 2013), Dublin Core[5], the Data Catalog Vocabulary[6], and Schema.org[7]. The model describes the six core entities, their attributes, and their relations. It has been visually represented using Graffoo[8] diagrams and formally defined as an OWL ontology, the **SKG-IF Ontology (SKG-O)**.

[4] http://www.sparontologies.net
[5] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[6] https://www.w3.org/ns/dcat#
[7] http://schema.org
[8] https://essepuntato.it/graffoo

SKG-O is composed of six modular ontologies, one for each core entity. The ontology and its submodules are available at the links reported in Table 1; more information can be found online at https://skg-if.github.io/data-model.

SKG-IF Ontology (SKG-O)	https://w3id.org/skg-if/ontology/
SKG-O: research-product	https://w3id.org/skg-if/ontology/research-product/
SKG-O: agent	https://w3id.org/skg-if/ontology/agent/
SKG-O: grant	https://w3id.org/skg-if/ontology/grant/
SKG-O: venue	https://w3id.org/skg-if/ontology/venue/
SKG-O: topic	https://w3id.org/skg-if/ontology/topic/
SKG-O: data-source	https://w3id.org/skg-if/ontology/data-source/

Table 1 - SKG-IF Ontology and submodules.

3.2. Interoperability framework

The interoperability framework is the heart of the SKG-IF and translates the data model by defining the actual metadata properties and format for each core entity. Key aspects include:

- **Local Identifiers:** Each entity must have a mandatory `local_identifier`, which is **always interpreted as a URL**. If a non-URL string is provided, it is combined with a **base URL** defined in the **JSON-LD preamble**[9]. The preamble specifies the `@context` pointing to the SKG-IF JSON-LD context and a `@base` URL following the pattern `https://w3id.org/skg-if/sandbox/<provider acronym>/`. Sources can also create on-the-fly identifiers[10] in case an entity type is not a first-class citizen in an SKG and therefore does not have a local identifier.
- **External Identifiers:** Entities can have a recommended `identifiers` property[11], which is a list of objects. Each object specifies the `scheme` (a string indicating the identifier type, e.g., `orcid`, `doi`) and the `value` of the external identifier.
- **Entity Type:** A mandatory `entity_type` field indicates the specific type of the entity being serialised. The allowed values are `agent`, `person`, `organisation`; `grant`; `venue`; `topic`; `datasource`; and `product`.
- **Properties:** Each entity type has a set of specific metadata properties to capture relevant aspects. For example, an `Agent` can have properties like `name`, `given_name`, `family_name`, and `affiliations`. A `Data source` can have `persistent_identity_systems` and `data_source_classification`. A `Grant` includes `grant_number`, `title`, and `funding_agency`. A `Research product` can have `titles`, `abstracts`, and `topics`. A `Venue` can have

[9] https://skg-if.github.io/interoperability-framework/#json-ld-preamble
[10] https://skg-if.github.io/interoperability-framework/#local-identifiers-of-entities
[11] https://skg-if.github.io/interoperability-framework/#external-identifiers-of-entities

`acronym` and `type`. A `Topic` has `labels`. Top-level properties are intentionally set with optional cardinality in order to enable data exchange even for a subset of information from highly specialised SKGs (e.g., providing only DOI to DOI citations).

For brevity, we preferred not to report here the exhaustive list of properties and their formats for each entity. Comprehensive information on the metadata properties can be found online under https://skg-if.github.io/interoperability-framework; the documentation pages describing the properties and their format for each SKG-IF core entity are reported in Table 2.

The SKG-IF data created is aligned with the SKG-IF Ontology via the SKG-IF **JSON-LD context**. Further details are available online at https://skg-if.github.io/context. The current version of this context is available online at https://w3id.org/skg-if/context/skg-if.json.

A SHACL document has also been developed for semantic validation of the data; this can be found online at https://w3id.org/skg-if/validation/shacl.

Research product	https://skg-if.github.io/interoperability-framework/docs/research-product.html
Agent	https://skg-if.github.io/interoperability-framework/docs/agent.html
Grant	https://skg-if.github.io/interoperability-framework/docs/grant.html
Venue	https://skg-if.github.io/interoperability-framework/docs/venue.html
Topic	https://skg-if.github.io/interoperability-framework/docs/topic.html
Data source	https://skg-if.github.io/interoperability-framework/docs/data-source.html

Table 2 - Documentation pages describing the properties of SKG-IF entities.

3.2 API Specification

The SKG-IF also includes an API specification to facilitate and technically render feasible the data exchange between SKG-IF adopters. The API specification is drawn using the OpenAPI standard[12] and is available only at https://w3id.org/skg-if/api/skg-if-openapi.yaml.

The specification provides dedicated endpoints for each SKG-IF core entity and supports querying and filtering. More information can be found in the online documentation at https://skg-if.github.io/api.

4. Extending the SKG-IF

The SKG-IF is designed to be **extensible** and address the evolving and diverse needs of scientific communities and projects. Introducing extensions to the SKG-IF serves as a strategic step towards addressing the evolving and diverse needs of scientific communities and projects. The concept of extensions within the SKG-IF is designed to cater to shared

[12] https://www.openapis.org

interests and requirements that the existing standard entities and properties may not fully meet.

By allowing for the development of extensions that enrich the core model entities of the SKG-IF, the framework can accommodate specialised use cases, specific research needs, and emerging data requirements without compromising the integrity and structure of the core framework. Extensions enable further development of the framework's capabilities without disrupting existing entities, thereby maintaining compatibility, coherence, and interoperability across diverse datasets and knowledge graphs.

Key participation guidelines for proposing an extension include:

- **Shared Interest/Need:** Extensions should address a collective interest or requirement identified within projects, communities, or domains rather than catering to individual or isolated needs. This principle ensures that extensions serve a broader purpose and contribute to the overall advancement of scientific knowledge representation.
- **Non-Interference:** Extensions should not interfere with or disrupt the entities and APIs already defined within the SKG-IF or other endorsed extensions. They should not serve as shortcuts for sharing information that should be placed elsewhere within the framework. By adhering to this principle, the integrity and coherence of the SKG-IF are preserved.

An SKG-IF extension can cover the following:

- **Core entity extensions:** add new properties and relations that extend the set of properties and relations of core entities to address the needs of the specific case statement;
- **Brand-new additions:** introduce brand-new entities that reflect semantics and structure different from the core entities, i.e., no core entity can be seen as a super-entity of community entities.
- **Additional API specifications:** extend the core SKG-IF API specifications to accommodate new properties and entities, as well as define new filters and functionalities for the RESTful endpoints.

To apply for an extension, proponents should open a **new issue** in the SKG-IF extensions repository on GitHub using the provided template. Requests are evaluated by the members of the RDA WG. Accepted extensions will have a new repository created where the essential documentation needs to be laid out orderly to be automatically integrated within the main SKG-IF website via GitHub Actions. Each extension should specify an **agreed namespace** to prevent clashing with other extensions.

The lifecycle of an extension falls outside the core WG's scope, but successful extensions may have aspects incorporated into the core model. Detailed information is available at https://skg-if.github.io/extensions.

At the time of writing, a number of extensions have been requested, and their current status and relevant information are reported in Table 3.

Acronym	Scope	Status	Link/request
RA-SKG	Indicators for products and agents	Development finalised	https://skg-if.github.io/ext-ra-skg/
SRV	New entity to model services	Under development	https://skg-if.github.io/ext-srv/
GEO-SKG	Geolocation and timestamps for products	Accepted	https://skg-if.github.io/ext-geo-skg/
SKG-PROV	Provenance tracking	To be discussed	https://github.com/skg-if/extensions/issues/4
RP-STDS	Research Products for Standards	To be discussed	https://github.com/skg-if/extensions/issues/5

Table 3 - Current status of SKG-IF extensions.

5. Contribution and Future Directions

The development of the SKG-IF is a **community effort**. If interested in participating in the discussion, there are two main options:

- Contribute to and consolidate the **SKG-IF core model** by subscribing to the Scientific Knowledge Graphs - Interoperability Framework RDA WG. At the time of writing, the WG meets twice a month and is dedicated to maintenance of the core model and discussion around extensions. More information available on activity section on the RDA website.
- Extend the SKG-IF to address specific requirements by reading the documentation about **extensions** and applying for a new one.

The WG envisions the SKG-IF playing a pivotal role in the **European Open Science Cloud (EOSC)** and, potentially, in a broader "**Global Open Science Cloud**". Future work may involve further refining the data model, enhancing the API specifications based on community feedback and extensions, and promoting the adoption of the SKG-IF within relevant initiatives.

At the time of writing, a few EU-funded projects expressed their interest in the SKG-IF and decided to adopt it within their scope. These are reported in Table 4. Furthermore, several initiatives and infrastructures have decided to adopt and map towards the SKG-IF. These are reported in Table 5.

6. Pointers to Online Documentation

The comprehensive online documentation for the SKG-IF is available at: https://skg-if.github.io. This documentation provides detailed information on the data model, API specification, extensions, JSON-LD context, examples, and other relevant aspects of the SKG-IF.

Project name	Grant ID	Website	Notes
SciLake	10.3030/101058573	https://scilake.eu	Using the model, but not the API specifications.
GraspOS	10.3030/101095129	https://graspos.eu	The project contributed to the development of the extension guidelines and successfully completed the development of the first SKG-IF extension.
OSTrails	10.3030/101130187	https://ostrails.eu	The project coordinated the development of the OpenAPI specifications.
LUMEN	10.3030/101187940	https://lumenproject.eu/	The project uses SKG-IF as one of the possible formats to share data.
GRAPHIA	10.3030/101188018	https://graphia-ssh.eu/	The project uses SKG-IF as one of the main mechanisms to enable federated queries and data sharing among the SKG involved.

Table 4 - Projects adopting SKG-IF.

Name	Website	Current status
OpenAIRE	https://www.openaire.eu	Under development
OpenCitations	https://opencitations.net	Under development
GoTriple	https://www.gotriple.eu	Under development
Novi-Sad University CRIS system	https://cris.uns.ac.rs/sr	Prototype delivered

Table 5 - Initiatives and infrastructures adopting SKG-IF.

References

- Amir Aryani, Martin Fenner, Paolo Manghi, Andrea Mannocci, & Markus Stocker. (2020, August 28). Open Science Graphs Must Interoperate!. Scientific Knowledge Graphs at the International Conference on Theory and Practice of Digital Libraries 2020 (TPDL) (SKG). https://doi.org/10.1007/978-3-030-55814-7_16

- Peroni, S., Shotton, D. (2018). The SPAR Ontologies. In: Vrandečić, D., et al. The Semantic Web – ISWC 2018. ISWC 2018. Lecture Notes in Computer Science(), vol 11137. Springer, Cham. https://doi.org/10.1007/978-3-030-00668-6_8
- Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. (2013). PROV-O: The PROV Ontology. (W3C Recommendation). World Wide Web Consortium. http://www.w3.org/TR/2013/REC-prov-o-20130430/