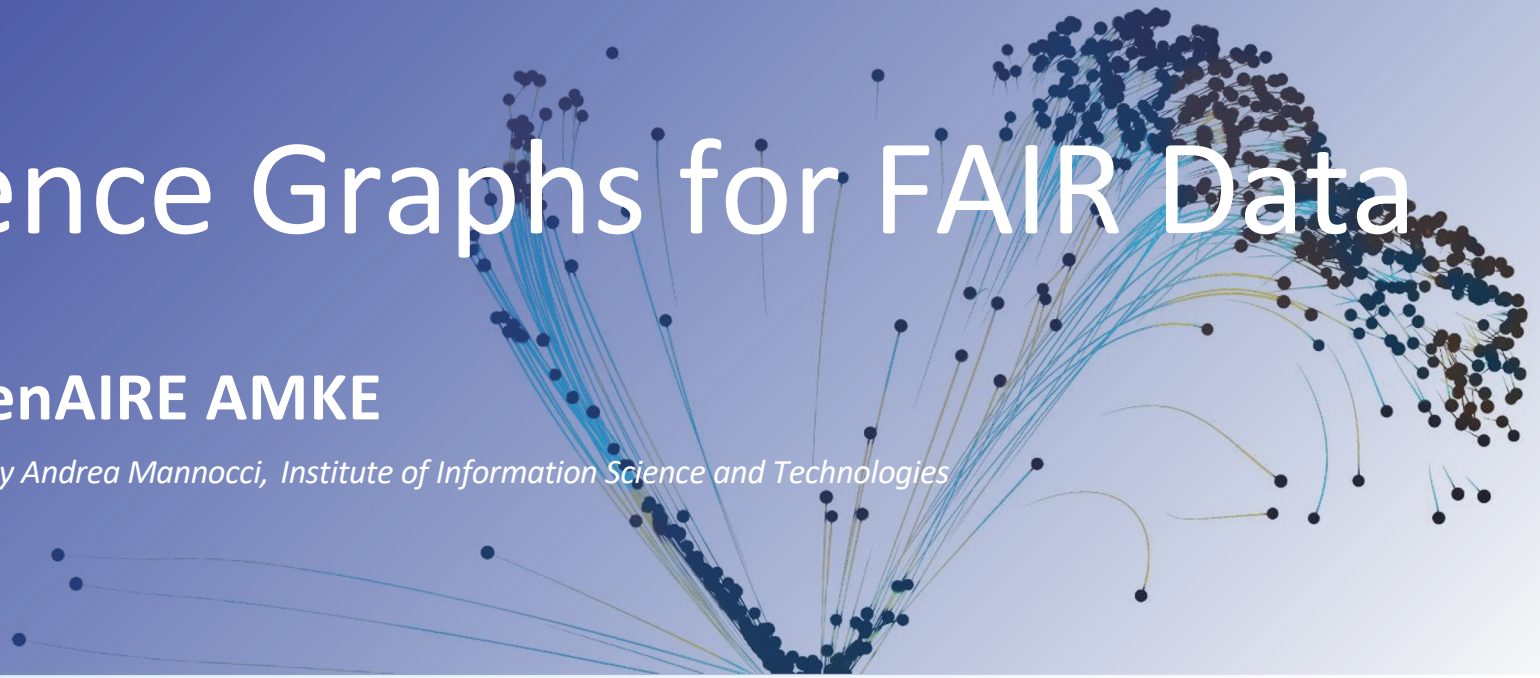
Open Science Graphs for FAIR Data

Paolo Manghi, OpenAIRE AMKE

Many of these slides are kindly shared by Andrea Mannocci, Institute of Information Science and Technologies



OpenAIRE

European Commission

@openaire_eu

Open Science Graphs: the origins

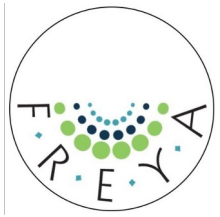
The **Open Science movement** is urging researchers to leave traces of their daily quest

- Deposition of **research entities** (“beyond the PDF”)
  - Research data, Software, Ideas & concepts, Tools & protocols, Methods & results, etc.
- With **relationships** between entities and to authoritative registries
  - Data (e.g. re3data), authors and authorship (e.g. ORCiD), organisations (e.g. ROR, GRID.ac), projects/funders (e.g. CORDIS), services, and so on.

Such information ends up in a **plethora of different services** scattered across the Web

Open Science Graphs: the rise

Many initiatives spawned so to contribute to and/or consume such information and build coherent, complete, and cleaned collections



Open Science Graphs interoperability

Clarivate
Web of Science™

Let's simplify data and knowledge exchange!

Scholexplorer

Scopus®

ResearchGraph

ORKG

Research.fi

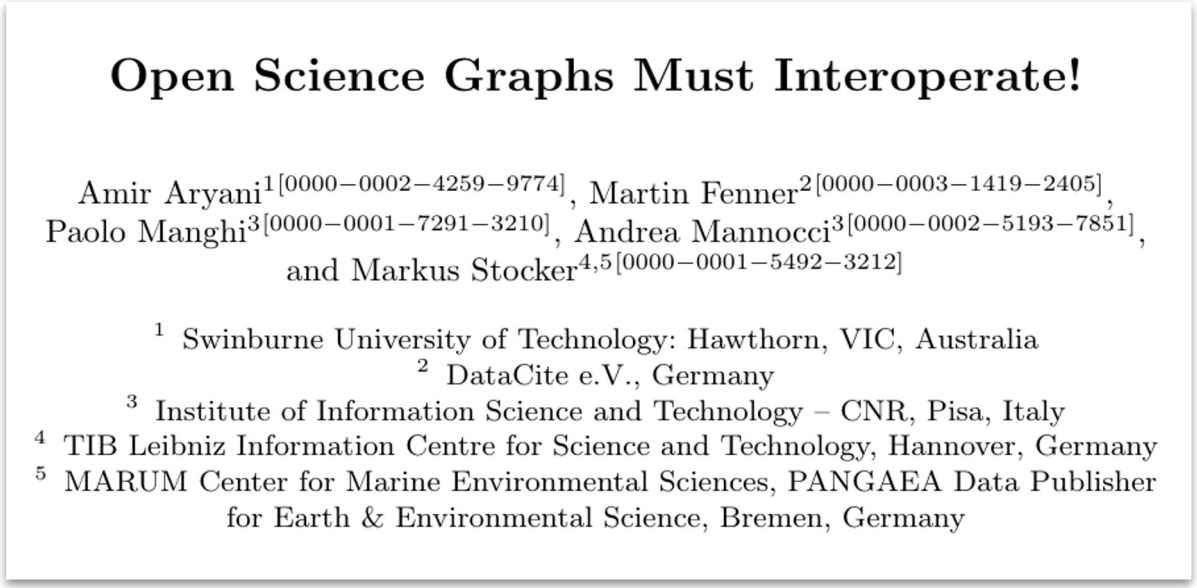
Crossref

OpenAlex

A first work in this direction

- A classification framework for OSGs
- The funding reasons so to foster an OSGs interoperability framework

https://zenodo.org/record/4006474



Open Science Graphs Must Interoperate!

Amir Aryani[1][0000−0002−4259−9774], Martin Fenner[2][0000−0003−1419−2405], Paolo Manghi[3][0000−0001−7291−3210], Andrea Mannocci[3][0000−0002−5193−7851], and Markus Stocker[4,5][0000−0001−5492−3212]

[1] Swinburne University of Technology: Hawthorn, VIC, Australia
[2] DataCite e.V., Germany
[3] Institute of Information Science and Technology – CNR, Pisa, Italy
[4] TIB Leibniz Information Centre for Science and Technology, Hannover, Germany
[5] MARUM Center for Marine Environmental Sciences, PANGAEA Data Publisher for Earth & Environmental Science, Bremen, Germany

A framework for OSGs interoperability

Objectives

- **Capitalise on synergies** and non-negligible efforts for content acquisition, integration, enrichment performed locally at OSGs.
- Be the **backbone** for exchanging open metadata in Open Science

This can be achieved by agreeing upon a **lingua franca** so to support information flow on two level of abstraction

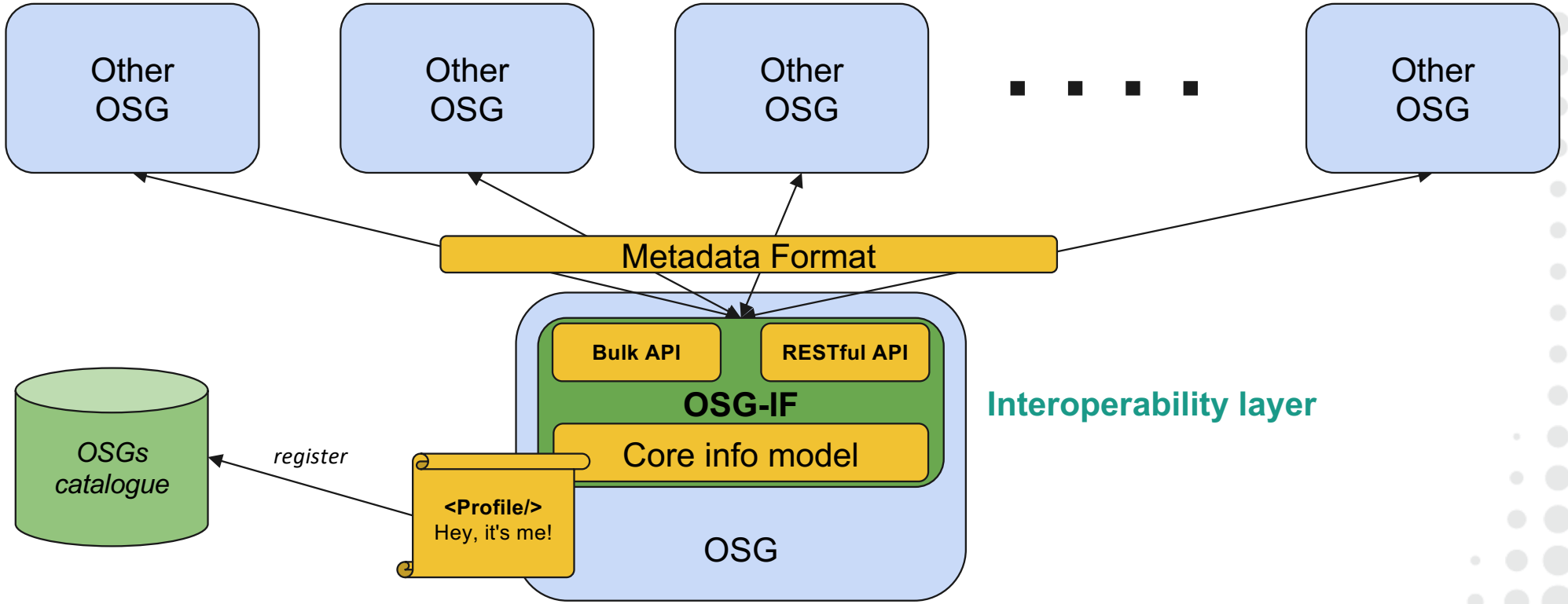
- **Information model** (maximise information exchange and flexibility)
- **Technological** (standards, exchange formats, primitives, APIs, etc.)

Scholix (http://www.scholix.org) is a successful example in this sense, despite targeting a much simpler scenario (i.e. literature-data linking).

Towards an interoperability framework for Open Science Graphs

- RDA IG on "OSGs for FAIR data" and related WGs (https://rd-alliance.org/groups/open-science-graphs-fair-data-ig)
- Graphs/initiatives available to experiment:
  - CERIF
  - Crossref
  - DataCite
  - DataCite PID Graph
  - OpenCitations
  - OpenAIRE Research Graph
  - OpenAlex
  - Open Research Knowledge Graph (ORKG)
  - ResearchGraph.org

High-level Architecture



Task Force 1 - OSG Core Information Model

Release 1.0, June - November 2022

1. Identification of the entities
2. Survey of (de facto) standard information models, PIDs, for the entities (e.g., CERIF, DataCite, JATS, Dublin Core, ROR.org)
3. Definition of the core model: incremental approach, introducing entities one by one

Task Force 2 - OSG Data Exchange Commons

Release 1.0, November 20th - January 30th

1. Survey on (de facto) Standard metadata formats for the entities (e.g. CERIF, DataCite, JATS, Dublin Core, ROR.org.) - *performed as part of the Core Information Model*
2. Definition/identification of Core Metadata Format
3. Crosswalks from known standards to Core Metadata Format (incremental, follows Information Model approach)

Task Force 3 - OSG Access Protocol Commons

Release 1.0, January - March 2023

1. Analysis of (de facto) standard protocols and approaches for exchanging Graphs data (e.g. RESTful API, GraphQL, SPARQL anything, etc.)
2. Identification of a set of protocols

Task Force 4 - OSG Profiles

Release 1.0, January - June 2023

1. Definition of OSG profile, attributes and vocabularies
2. Short specification document (published in Zenodo.org)

Outcome

OSG name	Added by/ Source	Entities: Research products (EOSC terminology): Literature: intended for reading by humans (article, thesis, peer-review, blog posts, books, reports, patents, etc.)	Research Data: self-contained, persistently identified digital assets intended for processing (e.g. files containing: tables, metadata collections, dumps; persistent	Research Software: (definition from RDA WG) Research Software includes source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose. Note that software components (e.g., operating systems, libraries,	Others: any digital asset, uniquely identified, whose nature does not fall in the first three types	Author (researcher)	Organizations (Institutions, companies, funders, publishers)	Journals	Topics (Ontologies, Taxonomies etc.)
OpenAIRE Research Graph		✔	✔	✔		✔	✔	✔	✔
European Open Science Cloud	Paolo Manghi	✔	✔	✔		✔	✔	✔	✔
OpenCitations	arcangelo.massari@unibo.it	✔	✔	✔	✔	✔	✘	✔	✘
OpenAlex	jason@ourresearch.org	✔	✔	✔		✔	✔	✔	✔
Dimensions	m.pasin@digital-science.com	✔	✔	✘	✔	✔	✔	✔	✔
SemanticScholar	Andrea Mannocci	✔	✘	✘	✘	✔	✘		✔
ResearchGraph	Amir Aryani	✔	✔	✔		✔	✔	✔	✔
Open Research Knowledge Graph	Markus Stocker	✔	✔	✔		✔	✘	✔	✔
Crossref	Martyn Rittman	✔	✔	✔		✔	✔	✔	✔
ORCID	Paolo Manghi	✔	✔	✔		✔	✔	✘	✘
Datacite	Julián Garrido. Compared with DataCite Metadata Schema V 4.4	✔	✔	✔		✔	✘	✔	✘
CERIF-CRIS	Jan Dvorak								
NARCIS - Netherlands	wim.hugo@dans.knaw.nl	✔	✔						
Cristin	martin.thomas.horsch@nmbu.no	✔							
DataverseNO	martin.thomas.horsch@nmbu.no		✔						
DaRUS	martin.thomas.horsch@nmbu.no		✔						
NFDI4Ing Terminology Service	martin.thomas.horsch@nmbu.no								
Google Knowledge Graph (datasets and other research outputs, entities)	noel.vizcaino@stfc.ac.uk	✔							

Modeled entities

- Products
- Authors
- Organizations
- Journals/Conferences
- Topics
- Projects

Properties and relationships per entity

ID	Property	Definition	Mandat., Recomm., Optional	Occ	Allowed values, examples, other constraints	Note	DCAT	OpenAlex	OpenAIRE	DataCite	CERIF	OpenCitations Data Model
1	*LocalIdentifier*	Unique code identifiying the product in the Graph	M	1			id (IRI) (note: e.g. any aliased @id in JSON-LD), top graph object		id			datacite:hasIdentifier
2	*Identifiers*		O	0..n			dcterms:identifier		pid or originalId			datacite:hasIdentifier
2.1	*identifierScheme*		M	1	vocabulary of identifier schemes?		use skos:Concept in dcterms:type for custom PID schemes		pid.scheme			datacite:usesIdentifierScheme
2.2	*identifierValue*		M	1	DOI, ArXiv,	We need to define/refer to existing vocabulary scheme	IRIs preferred but other PID/URLS accepted		pid.value			literal:hasLiteralValue
3	*Title*		M	1..n			dcterms:title		maintitle or subtitle			dcterms:title
4	*Abstract*		R	0..n			dcterms: description		description			
5	*Dates*		M	1..n			dcterms:PeriodOfTime { hasBeginning, hasEnd} Note: Use ISO 8061 timestamps with timezone		embargoenddate or publicationdate			prism:publicationDate
5.1	*Date*	Date format	M		2022, 31/12/2022							
5.2	*DateType*	Publishing date, Embargo Date	M									
6	*ResourceType*		M		Literature, Research Data, Research Software, Others				type			
6.1	*ResourceTypeGeneral*		O									
Relationships												
	toAffiliation(Author, Organization)	Reference to a pair <author, organization>, i.e. organization to which the author was affiliated when generating this product	Author M Organization O			how to model affiliations withpairwise relationships? do we need a separate entity?	For membership somethink akin to this (FOAF) { "@id": "http://example.org/#band-1", "@type": "mo:MusicGroup", "foaf:name": "Fugazi", "mo:activity_start": "1987-01-01T00:00:00Z", "foaf:member": [{ "@id": "http://example.org/#imk", "@type": "mo:MusicArtist",					:silvio-peroni a foaf:Person ; pro:holdsRoleInTime :silvio-peroni-at-unibo . :silvio-peroni-at-unibo a pro:RoleInTime ; pro:withRole scoro:affiliate ; pro:relatesToOrganization <https://ror.org/01111rn36> ; pro:relatesToDocument

Join our group!

Consiglio Nazionale delle Ricerche

ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"

Thank you!

Paolo Manghi, OpenAIRE AMKE & ISTI-CNR

paolo.manghi@isti.cnr.it
andrea.mannocci@isti.cnr.it