

EOSC Research Discovery Graph Service (RDGraph)

CRIS2024 pre-workshop: Hands-on Research Data, 14 May, 2024

Thanasis Vergoulis, OpenAIRE



Funded by
the European Union



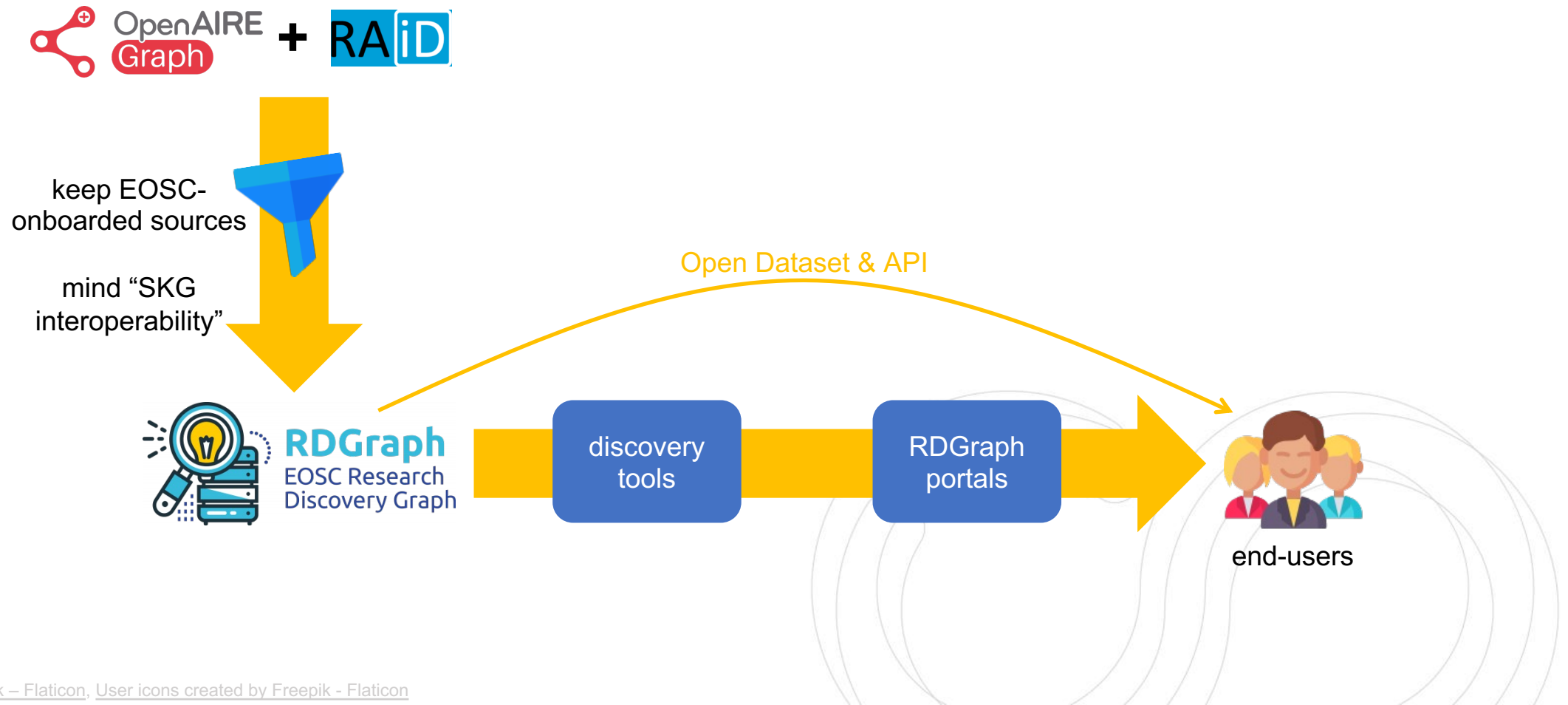
RDGraph: motivation

exponential growth of
scientific output

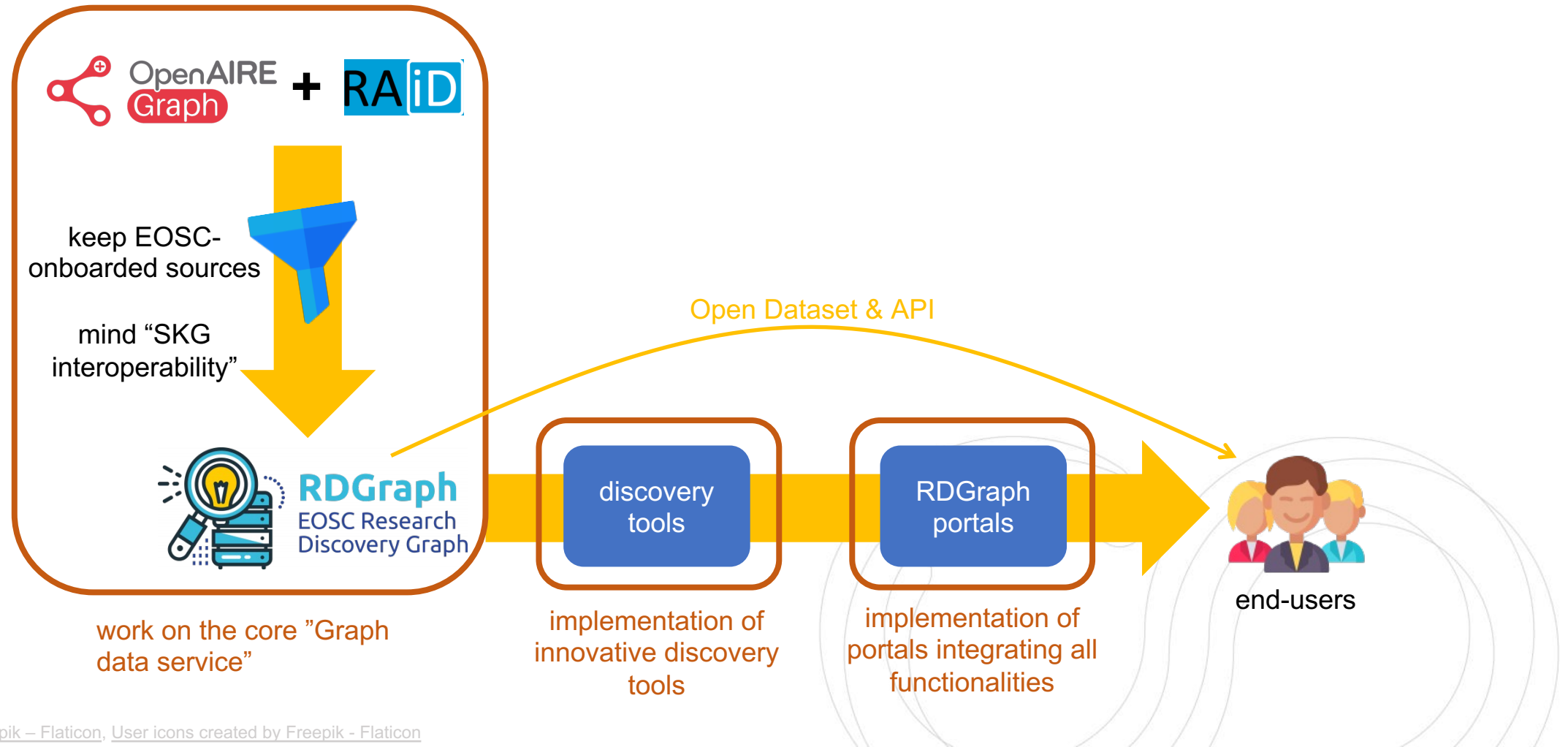
difficult to identify
really valuable research



RDGraph: the concept



RDGraph: the concept



Graph data service

What has been achieved

- Workflow for creating SKG-IF compliant RDGraph [1]
- Dataset soon to be deposited on Zenodo
 - Currently available on gDrive [2]
- Community-specific sub-datasets supporting the activities in community recommendations

Next steps

- Import of RAiDs (exploit **Datacite**)
- API to become more compliant to SKG-IF

[1] <https://skg-if.readthedocs.io/en/latest/products.html>

[2] https://drive.google.com/drive/folders/1mMuJPTAI346_OWqCBRsd05o_1krfOLAZ?usp=sharing



Natural language search

FC4EOSC NL Search

Search over the RD Graph with Natural Language

Show the full names of authors with more than 10 publications Search Examples -

	fullname
0	John Sandlos
1	Claudio Ciavatta
2	Sue Lin Ngan
3	Ana Andrés
4	Iftekhhar Hasan
5	Eran Stark
6	Soon-Bark Kwon
7	Donato Loddo
8	Wen Chi Hou
9	Alan J. D. Tennyson

Predicted SQL Query ^

```
SELECT t1.fullname FROM author AS t1 JOIN result_author AS t2 ON t1.id = t2.author_id GROUP BY t1.id HAVING COUNT(*) > 10
```

SQL Query Candidates ^

Natural language search

What has been achieved

- Creation of the NL Search component, using a Pre-trained Language Model, fine-tuned for Text-to-SQL
 - Fine-tuning done with the Spider multi-domain text-to-SQL dataset
- Creation of a complete pre/post-processing pipeline to optimize the use of the PLM
 - Handling and processing the necessary inputs of the model (question, DB schema, data)
 - Automatic correction and refinement techniques to improve the quality of generated SQL queries
- Connection of the component to a relational dump of the RDGraph, using sample communities
- Deployment of a demo and API to allow users and developers to access the service

Next steps

- Training specifically for the RD Graph: Improve performance for our use cases
- Update the RD Graph instance with data and communities more relevant to the project
- Integration to the RDGraph Portals



Impact-based search

The screenshot shows a search interface with a search bar containing 'semantic search'. Below the search bar, there are filters for 'RESEARCH PRODUCTS (14,997)', 'PROJECTS (40)', 'SERVICES (0)', and 'ORGANIZATIONS (0)'. The main results section is titled '14,997 Research Products for semantic search' and includes a 'DOWNLOAD RESULTS' button. A filter for 'OPEN ACCESS' is active. The 'Sort by' dropdown is set to 'Popularity'. The first result is 'Searching for MobileNetV3', a publication from 2019 by Andrew Howard et al. The impact metrics for this result are: Citations: 1,450; Popularity: Top 0.01%; Influence: Top 0.1%; Impulse: Top 0.01%. The result is powered by BIP! and has 1K citations.

semantic search

RESEARCH PRODUCTS (14,997) PROJECTS (40) SERVICES (0) ORGANIZATIONS (0)

14,997 Research Products for semantic search [DOWNLOAD RESULTS](#)

OPEN ACCESS x

Results per page: 10

Sort by: Popularity

1 2 3 4 5 >

Searching for MobileNetV3
Publication » Article, Preprint, Conference object • 2019 • IEEE
Authors: Andrew Howard; Ruoming Pang; Hartwig Adam; Quoc V. Le; +8 Authors
DOI: 10.1109/iccv.2019.00140, 10.48550/arxiv.1905.02244

We present the next generation of MobileNets based on a combination of complementary architecture design. MobileNetV3 is tuned to mobile phone CPUs through a combination of...

arXiv.org e-Print Ar... Share Cite

Citations 1,450
Popularity Top 0.01%
Influence Top 0.1%
Impulse Top 0.01%

POWERED BY BIP!

1K

Impact-based search

What has been achieved

- Impact indicators (influence, popularity, impulse, citation count) for all research products are computed within the content provision pipeline of the Graph
- Impact indicators can be used to rank the results of the keyword-based search in the RDGraph portal
 - E.g., Retrieve publications for the keywords “Semantic search” and sort the results based on their number of citations by descending order
- Impact indicators can be used to filter the results of the keyword-based search in the RDGraph API
 - E.g., Retrieve only publications in Top 1% based on their calculated popularity score

Next steps

- Combine keyword-based relevance with impact indicators to rank keyword search results



Community recommendation profiles

What has been achieved

Developed three RDGraph **services**



S1: Author-centric research product recommendations (training using RDGraph's citation edges)



S2: Research field-specific product recommendations (online learning via user feedback)



S3: Similarity-based research product recommendations (training using textual data & metadata of research products)

Next steps

- Integrate services into RDGraph's UI
- Enhance **S3**'s performance and memory efficiency for scaling to 10M+ resources
- **S4:** Develop a community-based research product recommendation service



RAiD inference service



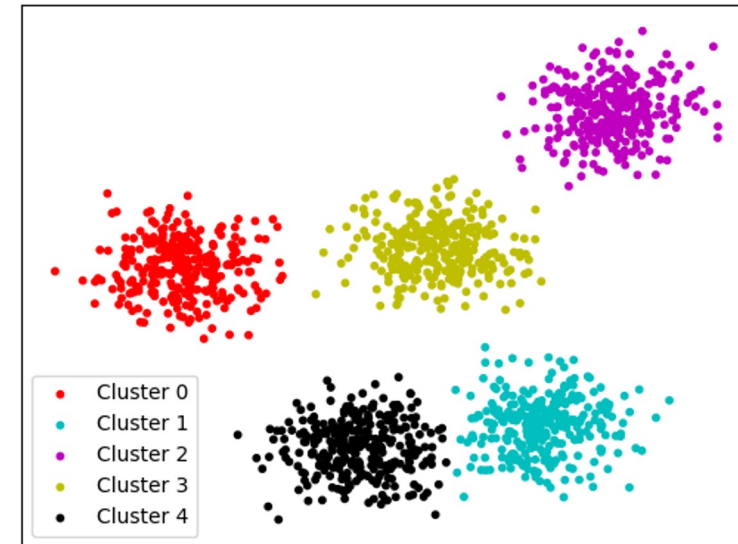
Metapath2Vec
DeepWalk

Including meaningful relationships for a RAiD

- research products versions
- institutions
- projects
- authors

Based on a 2-stage process

- stage 1: creation of research product N-dimensional embeddings
- stage 2: clustering of embeddings to create RAiDs



- Each cluster represents a RAiD

first batch of recommendations: by the end of June

RDGraph portals

HAL AMU View 2 versions Link to Share Cite Claim

Water-Enhanced Flux Changes under Dynamic Temperatures in the Vertical Vapor-Phase Diffusive Transport of Volatile Organic Compounds in Near-Surface Soil Environments

Publication » Article • 09 Jun 2021 • English • HAL CCSD • PeerJ Inc. (Copyright policy) • EC

Authors: Monami Kondo, Noriaki Watanabe, Kengo Nakamura, Mizuki Yamada, Jiajie Wang, Takeshi Komai
 DOI: [10.7717/peerj.1150](https://doi.org/10.7717/peerj.1150)

Summary

References (53)

Metrics

Recommended

Abstract

International audience It is commonly believed that vision is impaired during saccadic eye movements. However, here we report that some visual stimuli are clearly visible during saccades, and trigger a constriction of the eye's pupil. Participants viewed sinusoid gratings that changed polarity 150 times per second (every 6.67 ms). At this rate of flicker, the gratings were perceived as homogeneous surfaces while participants fixated. However, the flickering gratings contained ambiguous motion: rightward and leftward motion for vertical gratings; upward and downward motion for horizontal gratings. When...

[View more >](#)

Keywords

Vision science, Intrasaccadic perception, [SCCO.PSYC]Cognitive science/Psychology, General Neuroscience, [SCCO.PSYC] Cognitive science/Psychology, Saccadic suppression, Neuroscience, Eye movements, Medicine, Experimental psychology, General Agricultural and Biological Sciences, General Medicine, R, Psychiatry and Psychology, Pupillometry, General Biochemistry, Genetics and Molecular Biology

Powered by OpenAIRE Research Graph. Last update of records in OpenAIRE: Nov 20, 2019
[See an issue?](#) [Give us feedback.](#)

Citations	16
Popularity	TOP 1%
Influence	TOP 1%
Downloads	200
Views	500

SDGs View all & suggest >

- 1 No Poverty
- 10 Reduced Inequalities

Fields of Science (6) View all & suggest >

- Engineering
 - Electrical and Electronic Engineering
 - Mechanical Engineering
- Natural sciences

Funded by View all >

European Commission | ACTPUP, European Commission | OPENAIRE, FWF| Turbulence in the Edge of Magnetised Plasmas: Emergent Structures and Transport, EC| EUROfusion

Related to Research Communities View all >

- [Enermaps](#)
- [Digital Humanities & C...](#)
- [Lorem Ipsum](#)

Natural search, real results.

Search as you speak. Intuitive, natural, effortless.

How many publications concern COVID-19 and are related to OpenAIRE? ✕ 🔍

Note: Each new question deletes the previous session

Bot
 Lorem ipsum dolor sit amet, consetetur sadipscing elitr. Lorem ipsum dolor sit amet, consetetur sadipscing elitr.

Recommendations per related community

Enermaps

Digital Humanities & C...

Lorem Ipsum

Lorem Ipsum

[VIEW ALL \(41\) →](#)

Thank you!

thanasis.vergoulis@openaire.eu

Also do not miss the CRIS2024 Tutorial titled “**Facilitating the discovery and identification of EOSC resources**” (Thursday 10:30am-12:00am CEST)

