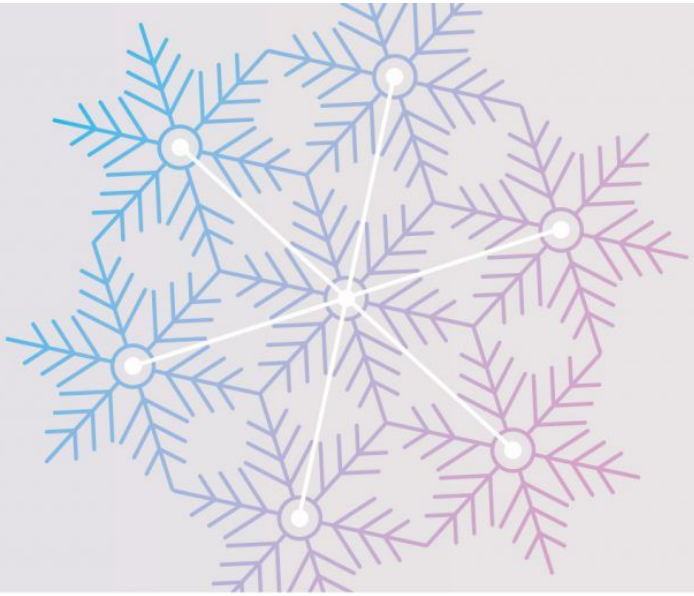
eosc

Winter School 2025

20 - 23 January 2025 / Seville, Spain



eosc.eu | #eoscwinterschool2025

eosc | Focus

Funded by the European Union

Strategic Pillar 2: Contributing to the Web of FAIR Data and Uptake of AI

Gorka Epelde
OAEG1/TF4:

Wednesday, 22 Jan 2025

Funded by the European Union

Contribution to SP2 Priorities

- Which specific SP priorities or topics have been discussed?
  - 4 sessions: *"Resources for PID implementation support", "Assessment mandate authorities for PID & LTDR policies", "AI-ready PIDs", "PIDs - LTDR for sensitive data"*
  - Scene: PIDs fundamental for EOSC Federation ⇒ Need of assessment mandate authority & resources for PID implementation in Nodes
  - Main results:
    - **Supporting** PID **implementation**: Tools are coming (FC4E & F-I) (CAT, PID Knowledge Base, …)
    - Suggestion for **governance**: EOSC Federation adopts the PID Policy (so all policies are together)
      - Some fall-back options discussed, but none considered preferable (e.g. global authority)
    - **PIDs for AI:** more efficient (incl. energy demand) and precise use of AI, PIDs are essential for this (e.g. FDOs)
    - **AI for PIDs**: no time for detailed discussion (e.g. on metadata enhancement) (⇒ other OAEGs&TFs?)
    - Access to **sensitive metadata** should be managed by **PID Managers** (i.e. at repository level)
- How do these topics align with mid- & long-term objectives defined during the WS?
  - Governance & orchestration of technologies / policies for a web of FAIR data and uptake of AI
- How do OAEG & TF contribute to advancing these priorities? (Highlight specific actions, initiatives, or outputs)
  - see above + OA1 & TF4 could be a good source of volunteers for peer reviewing of assessments

Funded by
the European Union

Challenges, Gaps, Recommendations and Next Steps

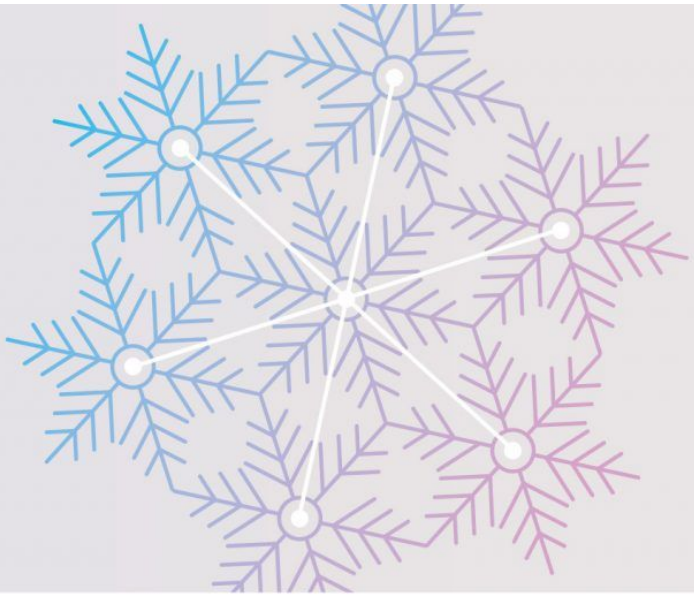
- Governance challenges
  - In order to build well-grounded Web of FAIR Data and enable uptake of AI, an **assessment body** is needed for PID & LTDR
- Technical challenges and alignment
  - Advanced **technologies, missing governance** and policies are missing (e.g., Data Type Registry, …)
  - Efficient and precise use of **AI for research purposes** requires PIDs and scientific knowledge graphs to serve as a receptacle for classifications, entity recognition, imputed relationships and annotated content, and to supplement requests via RAG-like enhancements
- Recommendations
  - EOSC PID Policy is fundamental for the federation and should be considered **together with other policies** for adoption.
  - PIDs mentioned at national/institutional level @MAR, but consider **EU-level recommendation** in the future EOSC documents (e.g., SRIA2.0)
- Next steps
  - Understand the governance options better (include the PID Policy)
  - Check the outcomes of the projects (will provide support for PID implementation and uptake of AI)

Funded by
the European Union

eosc

Winter School 2025

20 - 23 January 2025 / Seville, Spain



eosc.eu | #eoscwinterschool2025

eosc | Focus

Funded by the European Union

Strategic Pillar 2: Contributing to the Web of FAIR Data and Uptake of AI

Esteban Gonzalez

OA2/TF1+TF3: Interoperability

Wednesday, 22 Jan 2025

Funded by the European Union

Contribution to SP2 Priorities

- B. Accelerate the **adoption of interoperability and semantic artefact catalogues**. Support **AI-enabled services**, engaging discipline-specific groups that facilitate interoperability and reuse more automatically through automatic annotation, data linkage, data homogenization, data transformation, or other similar approaches, including global alignment.
- E. Use European and **domain-specific semantic artefact catalogues** in national infrastructures and guidelines aligned with European **standards and vocabularies**.
- F. Support adoption of general and **domain-specific standards** to increase adoption of FAIR practices and develop plans to facilitate reuse.
- I. Promote collaborations between national **infrastructures** and cross-border initiatives to promote consistent FAIR data standards and enhanced international **interoperability**.

Funded by
the European Union

Alignment with WS 2024 and OA/TF contribution

Mid-term objectives:

- To frame inter-project collaboration that allows seamless onboarding of future projects
- Contribute to shaping the vision of SRIA 2.0

Long-term objectives:

- The significance of this event is underscored by its potential to bring together a diverse range of perspectives, offering a rich blend of insights and experiences.

Contributions of Opportunity Area Expert Groups (OAEGs) & Task Forces (TFs):

- Identification of use cases to use AI to solve interoperability problems.
- Identification of use cases for each SP2 priority

Funded by
the European Union

Challenges and Gaps

- Two approaches: FAIR4AI and AI4FAIR
- To drive uptake, we need to show value - but **how to show value of AI**?
- How can semantic artefacts and their catalogues support the development of **AI-ready datasets**? What information we need to include?
- Integration of semantic artefact catalogues in research infrastructures by **using existing solutions in EOSC projects**.

Funded by
the European Union

Challenges and Gaps

- How can we **evaluate AI models to solve interoperability problems** such as crosswalks?
- **Explainability of AI models** and results is a challenge. How important is it? -> Combine with LLM with KG can solve the problems of hallucination
- Incentivization of researchers to use semantic artifacts.
- We need high volume of high quality data to train and to evaluate models **(“garbage in, garbage out”)**.
- Task expertise is essential to use AI effectively (e.g., writing code in ChatGPT) - you have to check and correct the output. **Results need supervision**

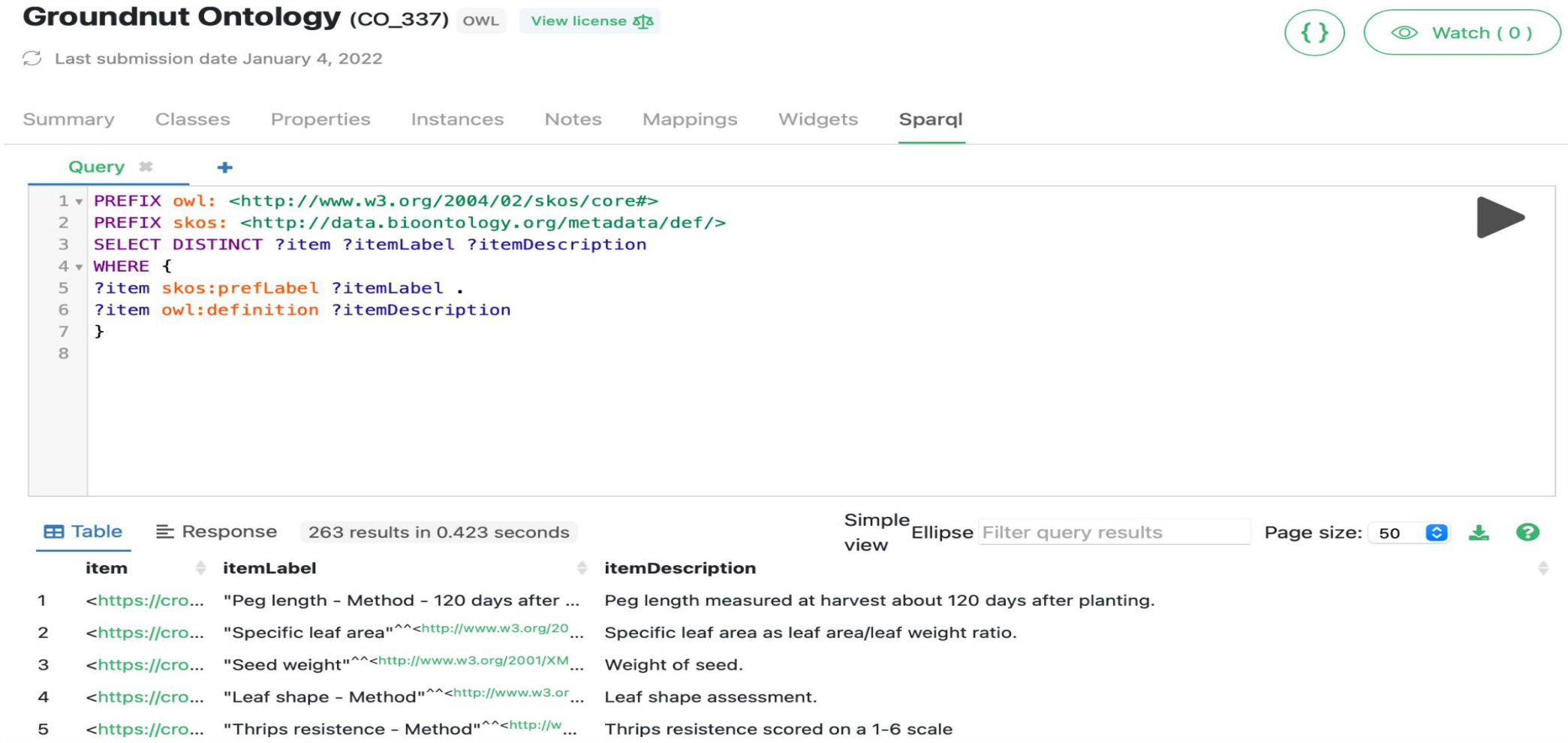
Funded by
the European Union

Recommendations and Next Steps

- Build benchmarks to evaluate models.
- Focus (also) on metadata, metadata curation/validation and coherent FAIR approaches to ensure quality of training and evaluation data.
- Theoretical discussions vs. pragmatic solutions in order to keep up with the U.S.
- How is AI already used in science? We need to collect use cases, caveats and “don’t use” cases.
- Ongoing developments to track:
  - FAIR for AI (knowledge graphs, use of semantic artefacts) and AI for FAIR (metadata quality assessment using LLMs) - both need to be happening.
  - Discussing what is “FAIR enough” for AI.

Funded by
the European Union

Ontoportal integration with LLM-KG

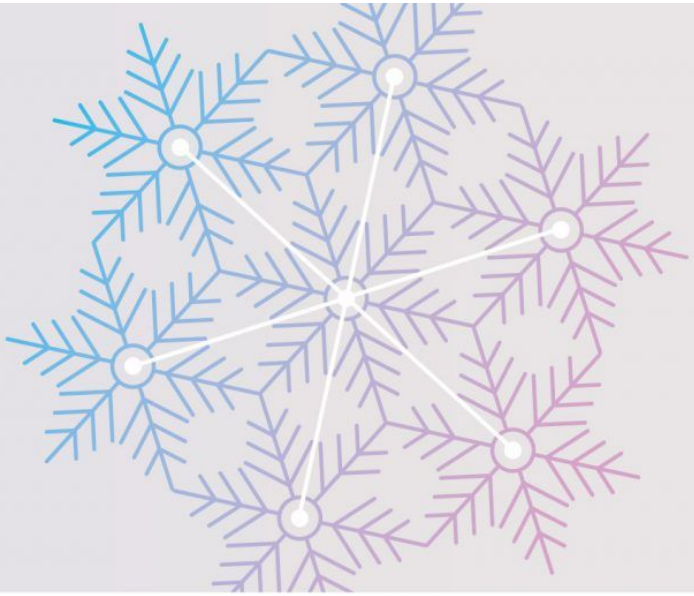


Funded by
the European Union

eosc

Winter School 2025

20 - 23 January 2025 / Seville, Spain



eosc.eu | #eoscwinterschool2025

eosc | Focus

Funded by the European Union

Strategic Pillar 2: Contributing to the web of FAIR data and the uptake of AI

Daniel Garijo
OA3/TF2/TF4

Wednesday, 22 Jan 2025 AM

Funded by the European Union

Contribution to SP2 Priorities

- Which specific SP priorities or topics have been discussed?
  - Is FAIR data enough for training AI models?
  - Can FAIR help making data AI ready?
  - Legal and Accessibility issues of datasets for AI models
  - Can AI help searching FAIR data?
  - AI for enhancing Digital Object metadata
- How do these topics align with the mid-term and long-term objectives defined during the WS 2024
  - it is a continuous progress, regarding requirements on data quality & FAIR data
- In what ways does the Opportunity Area Expert Group (OAEG) & Task Force (TF) contribute to advancing these priorities? (Highlight specific actions, initiatives, or outputs)
  - Discussing metadata accessibility issues (e.g., signposting)
  - Discussing ties with international initiatives (e.g., FAIR4ML)
  - Making clear where FAIR starts and ends (quality vs reproducibility vs FAIRness)

Funded by
the European Union

Challenges, Gaps, Recommendations and Next Steps

Challenge: Is FAIR data enough for training AI models?

- AI needs quality, and quality != fairness. But FAIR is key for **transparency** (provenance, reusability, etc.)
- Recommendations
  - Clarify the role of FAIR in AI (quality vs reproducibility vs reusability)
  - Training
- Next steps
  - Projects like **Eden** do quality but not specific to AI

Funded by
the European Union

Challenges, Gaps, Recommendations and Next Steps

Challenge: Legal and Accessibility issues of datasets for AI models:

- How to harmonize access conditions for letting agents what to use
- Recommendations
  - FAIR should help using the right data for the right purpose.
    - Risks/misuse of data for an AI model
- Next steps / relevant projects
  - Need an ODRL profile for dealing with AI data specific conditions
  - EOSC Entrust
  - FAIRCORE4EOSC and RO-Crate now support 5safes concepts for sensitive data
  - RDA working group on AI and data visitation

Funded by
the European Union

Challenges, Gaps, Recommendations and Next Steps

Challenge: Can FAIR help making data AI ready? (e.g. agents)

- Need to define what AI-ready means
  - AI-ready often requires cleaning and integration, which FAIR does not address
- Recommendations
  - Extend FAIRness for AI to address minimum metadata:
    - Completeness of AI datasets (missing values, imputation strategies, biases)
    - Variable description
    - Representation format
  - ML models should be FAIR! (and with transparent provenance)
- Next steps / relevant projects
  - Croissant data standard adoption
  - FAIR4ML and **crosswalks** effort
  - AI4EOSC?

Funded by the European Union

Challenges, Gaps, Recommendations and Next Steps

Challenge: Can AI help searching FAIR data?

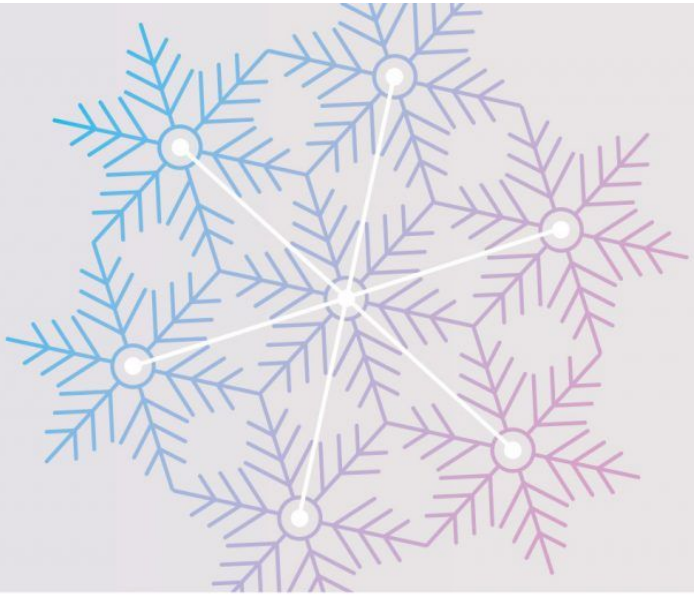
- How to incorporate AI as a curator of metadata
  - Keywords / categories
- Recommendations
  - Need to think about the role of AI in steering data spaces mappings (eg keyword/taxonomy based)
  - Need to assess AI as a recommender for selecting the right taxonomies (if they exist)
- Next steps / relevant projects
  - AI as a means to improve existing metadata (e.g., descriptions)
  - AI as a flagging system that needs human intervention
  - FAIRCORE4EOSC has MSCR, which could be used as a foundation for AI to populate mappings

Funded by the European Union

eosc

Winter School 2025

20 - 23 January 2025 / Seville, Spain



eosc.eu | #eoscwinterschool2025

eosc | Focus

Funded by the European Union

Strategic Pillar 2: Contributing to the Web of FAIR Data and Uptake of AI

Amanda Calatrava, Björn Grüning
OA4

Wednesday, 22 Jan 2025

Funded by the European Union

Contribution to SP2 Priorities

- Which specific SP priorities or topics have been discussed?
  - Model distribution - BYOD
  - Model training
  - Model inference + composition of AI pipelines
  - LLMs integrated into data analysis workbenches
- How do these topics align with the mid-term and long-term objectives defined during the WS 2024
  - Long-tail of science is able to use ML/AI models
  - RO-Crate export is here - including export to Zenodo
- In what ways does the Opportunity Area Expert Group (OAEG) & Task Force (TF) contribute to advancing these priorities? (Highlight specific actions, initiatives, or outputs)
  - Hackathons across the EOSC consortia and/or projects
  - Open-source tools and platforms to train and distribute (inference) AI models.

Funded by
the European Union

Challenges, Gaps, Recommendations and Next Steps

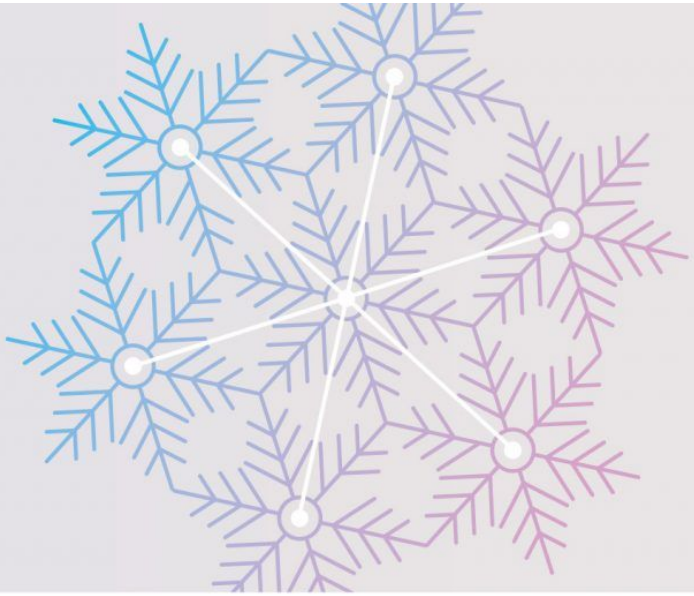
- Technical challenges and alignment
  - Users credential to access data between cloud storage systems used in the EOSC EU Node and AI4EOSC platforms (webdav)
  - Scaling … be prepared for success.
    - at some point we need to get back to “strace” again
    - Software quality varies between 3500 tools … who is debugging this, anyone OA7 ;)
- Recommendations / Questions
  - What is the necessity of LLM in EOSC? Can we / should we offer data-private LLM services as part of EOSC.
- Next steps
  - Explore further integration between platforms including the EOSC EU Node to increase the FAIRness

Funded by
the European Union

eosc

Winter School 2025

20 - 23 January 2025 / Seville, Spain



eosc.eu | #eoscwinterschool2025

eosc | Focus

Funded by the European Union

Strategic Pillar 2: Contributing to the Web of FAIR Data and Uptake of AI

Helen Clare
OA5

Tuesday, 22 Jan 2025

Funded by
the European Union

Contribution to SP2 Priorities

- Which specific SP priorities or topics have been discussed? (Web of FAIR data and AI adoption)
    - Rewards & Recognition & Adoption (J, K)
    - Competence Centres (B, E, G)
    - Engagement & Strategy (A-J)
    - Training Catalogue (B, E)
    - Curriculum/Learning Paths/Training Materials (G, B, E)
- For each topic
    - Discuss
    - Establish/Map with SP2 priorities
    - Define priority actions for OA5EG for 2025
- How do these topics align with the mid-term and long-term objectives defined during the WS 2024
- In what ways does the Opportunity Area Expert Group (OAEG) & Task Force (TF) contribute to advancing these priorities? (Highlight specific actions, initiatives, or outputs)
- Challenges, gaps, recommendations and next steps

Funded by
the European Union

Rewards & Recognition & Adoption (J, K)

Catalog and analyze emerging recognize/reward projects/pilots/practices that aim to incentivise adoption of EOSC (Open Science)

Action Plan by the Commission to implement the ten commitments of the Agreement on Reforming Research Assessment (ARRA)

Funded by
the European Union

Competence Centres (B, E, G)

- Bring parties together and create an overview of the various CC definitions/approaches taking the emerging EOSC nodes into consideration
- Consider organising a workshop or meeting between OA5EG, OSCARS and Skills4EOSC
- Build on Skills4EOSC output:
  - Attend next Skills4EOSC CC Net webinar 06/02/2025
  - A paper is in the making about CC
- Note: Also check out the Euro HPC National Competence Centers - long-established initiatives - for some countries this is the equivalent to a CC:
  - https://eurohpc-ju.europa.eu/national-competence-centres-high-performance-computing_en

Funded by
the European Union

Engagement & Strategy (A-J)

- Analysis of the EOSC Federation Handbook recommendations about Engagement, Skills and Training
- Define the scope of the workstream, considering past and existing engagement activities and the gaps to be addressed and complement with other engagement activities (what is the value of our OA5EG workstream… how does the work makes a difference)
- Identify the topics for HE Workprogramme 2025 EOSC calls focusing training and engagement to understand EC priorities

Funded by
the European Union

Training Catalogue (B, E)

- Build on EOSC-Future work about the Training Catalogue
- Define the vision for the EOSC Training Catalogue(s): what is needed, for whom, and which route are they following to get there
- Reach out to the EOSC EU Node via EC to explore current possibilities to add Training Resources
- Discuss with the Engagement and Strategy workstream the strategies for engagement with organisations and researchers needed to raise awareness about the opportunities in the EOSC Node

Funded by
the European Union

Curriculum/Learning Paths/Training Materials (G, B, E)

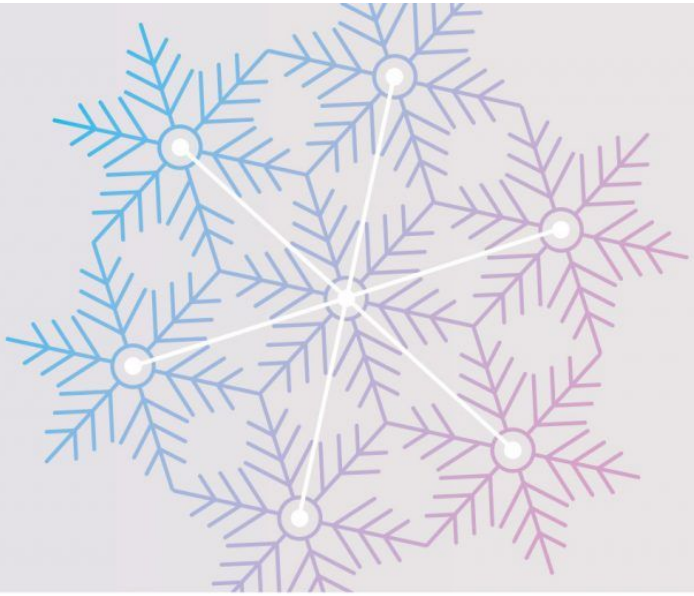
- Align existing methods for creating Learning Paths and map them
- OA5EG can learn from the cross-border training initiatives in the RIs (including Australia)
- To check with Skills4EOSC the sustainability plans
- Find out how much of CPD for data stewards is ongoing in the countries and institutes
- Connect with the RDA group
- Investigate how AI can extract metadata from training resources, preserving the data quality?

Funded by
the European Union

eosc

Winter School 2025

20 - 23 January 2025 / Seville, Spain



eosc.eu | #eoscwinterschool2025

eosc | Focus

Funded by the European Union

Strategic Pillar 2: Contributing to the web of FAIR data and the uptake of AI

Jiří Marek/Elli Papadopoulou
OA6/TF1, TF2

Wednesday, 22 Jan 2025

Funded by the European Union

Contribution to SP2 Priorities

Technical and Semantic Interoperability in OSC

- Which specific SP priorities or topics have been discussed?
    - SP2(F): Support adoption of both general and domain-specific standards to increase adoption of FAIR practices and develop plans to facilitate reuse (Objective 2).
    - SP2(B): Accelerate the adoption of interoperability and semantic artefact catalogues.

- In what ways does the OAEG6 & TF 1 contribute to advancing these priorities? (Highlight specific actions, initiatives, or outputs)
    - **We should discuss the machine interoperability of publication-data-supplementary data in more detail and learn from each other (publication/publisher/data stewards/research data worlds)**

Funded by
the European Union

Challenges, Gaps, Recommendations and Next Steps

- Technical challenges and alignment
  - How to effectively use metadata to describe a link between publication-research data-supplementary data + data availability statement
- Recommendations
  - European Diamond Capacity Hub (built in DIAMAS and CRAFT-OA projects) should became a thematic node of EOSC Federation in near future as a “voice of publication within the federation”
  - OA6 should play an important role for enabling discussion about publications within EOSC. EOSC is here for all Digital Objects (publication, data and software) - similar function as OA7 is doing for research software
  - Data dissemination and publishing criteria should be part of data stewardship courses. Some light weight training in repository and library publishing would really help to engage more in EOSC and the research data domain
- Next steps
  - Discussion between OAEG6 and EOSC EU Node on how to connect Diamond Discovery Hub to EOSC EU Node Resource Hub
  - Discuss semantic interoperability issues between publications and research data at the TF1 Semantic Interoperability subgroup meeting.

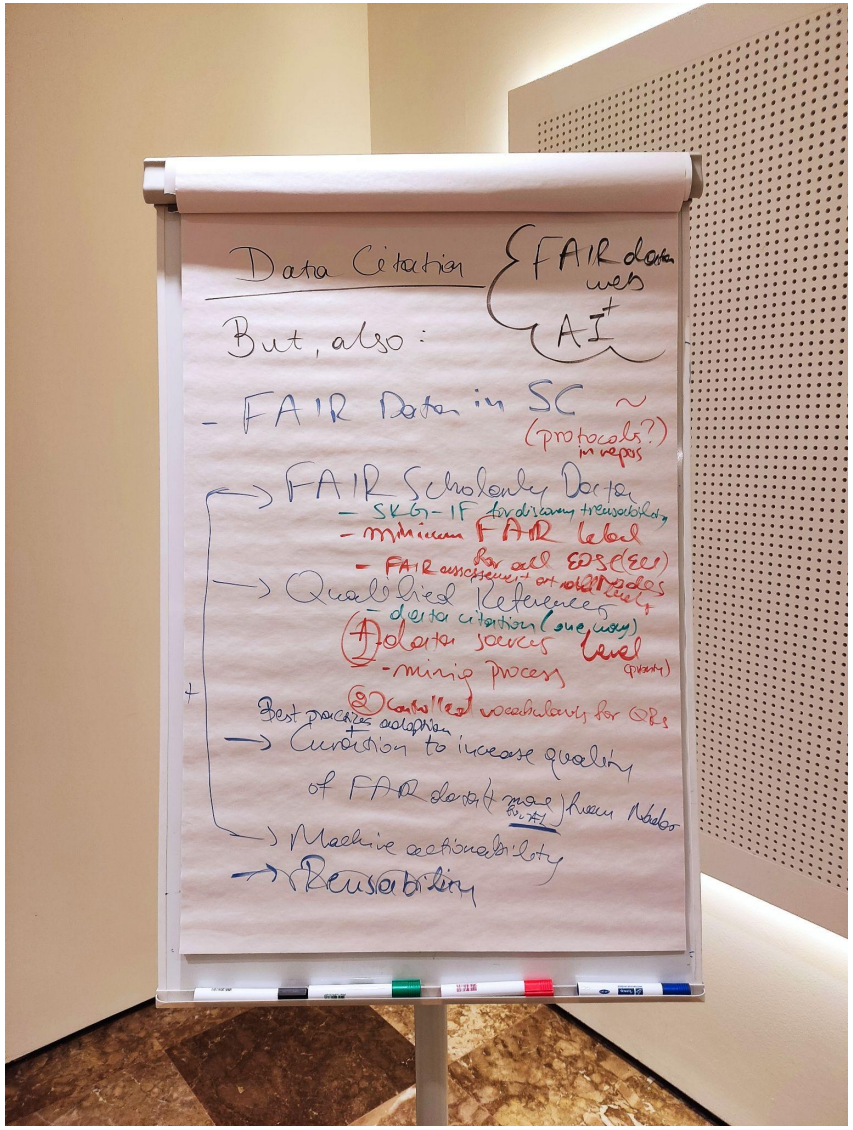
Funded by
the European Union

Contribution to SP2 Priorities

Machine-actionability within OSC - SKG-IF

- Which specific SP priorities or topics have been discussed?
  - SP2 (B): Interoperability of different FAIR objects communicated as open research outcome: SKG (scholarly knowledge graphs) should be able to get the interoperable metadata
  - SP2 (D): There are domain specific guidelines and standards, we need to establish the commons

- In what ways does the OAEG6 & TF1 & TF2 contribute to advancing these priorities? (Highlight specific actions, initiatives, or outputs)
  - **Open scholarly communication goes beyond the publishing process, other objects are more complicated to control. We need to work on standards and guidelines together with the community.**



Funded by
the European Union