Task Force Charter[1]
Semantic Interoperability

[1] Authors: Kurt Baumann (SWITCH); Oscar Corcho (Universidad Politecnica de Madrid); Martin Thomas Horsch (HLRS); Thomas Jouneau (UL); Marco Molinaro (INAF & ICDI); Silvio Peorni (University of Bologna); Andrea Scharnhorst (DANS-KNAW); Sadia Vancauwenbergh (ECOOM-UHasselt, euroCRIS); Lars Vogt (TIB) (in alphabetical order).
The TF coordinators opened the document to all which had declared an interest. Listed authors are those who attended the four preparatory meetings (May/June), and/or contributed to the text.
As the charter serves as a basis to solicit further participation also from communities not yet closely involved, we decided to add an introduction to the topic.

Introduction

EOSC and semantic interoperability

The European Open Science Cloud (EOSC) started as a European Commission initiative, and is (since 2020) a legal entity, an association. This association brings together "all relevant stakeholders to co-design and deploy a **European Research Data Commons** where data are findable, accessible, interoperable and reusable (FAIR), and also as open as possible".[2] In short, its goal is to develop a federated data infrastructure providing its users with services promoting Open Science practices.

EOSC aims: (1) to increase the value of scientific data assets by making them easily available to a larger number of researchers, across disciplines (interdisciplinarity) and borders (EU added value) and (2) to reduce the costs of scientific data management, while (3) ensuring adequate protection of information/personal data according to applicable EU rules and associate countries.

The vision of a Europe-wide shared data infrastructure is rooted in the ongoing digitization of empirical material, the further development of digital methods penetrating all fields of academia; and most fervently the technological possibilities offered by the Internet. That means, the ability to access resources and to execute analysis in a distributed networked way using the programmable web.

It is this way of networked exchange of information and collaboration which requires a specific discussion on issues of interoperability and more specifically semantic interoperability. To be able to use machines to connect to and process different data sources, standards defined under an open community management need to be applied so that those sources become more interoperable. Openness of data[3] is a prerequisite for both interoperability and the ability to share data.

In recent years, criteria have been formulated - the FAIR principles (Wilkinson et al. 2015) - to enable 'data fluidity' under the Open Science paradigm. **FAIR principles** and more specifically the role of **interoperability** form the foundation of the Task Force Semantic Interoperability.[4]

Semantic interoperability is a specific aspect of the pluriform concept of interoperability. "It ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words 'what is sent is what is understood'". (Corcho et al. 2021, p.20)

This charter starts from the EOSC Interoperability Framework (Corcho et al. 2021) as its manifesto (abbreviated EOSC-IF), and looks into ways to further foster discourse, knowledge exchange and the alignment of ongoing, concrete explorations related to semantic

---

[2] https://www.eosc.eu
[3] As defined in https://opendefinition.org/od/2.1/en/
[4] There are natural dependencies with other EOSC TF's (see section below)

interoperability. In the EOSC-IF, semantic interoperability is addressed as one layer in an interoperability framework: next to technical, organisational and legal interoperability.

Challenges to semantic interoperability

Having set out the wider boundary conditions under which this Task Force operates, it is important to flag out some of the intrinsic uncertainties, complexities, and challenges which accompany this work. Those have been identified in the section ‘Problems and Needs’ (EOSC IF, p. 16ff)[5]

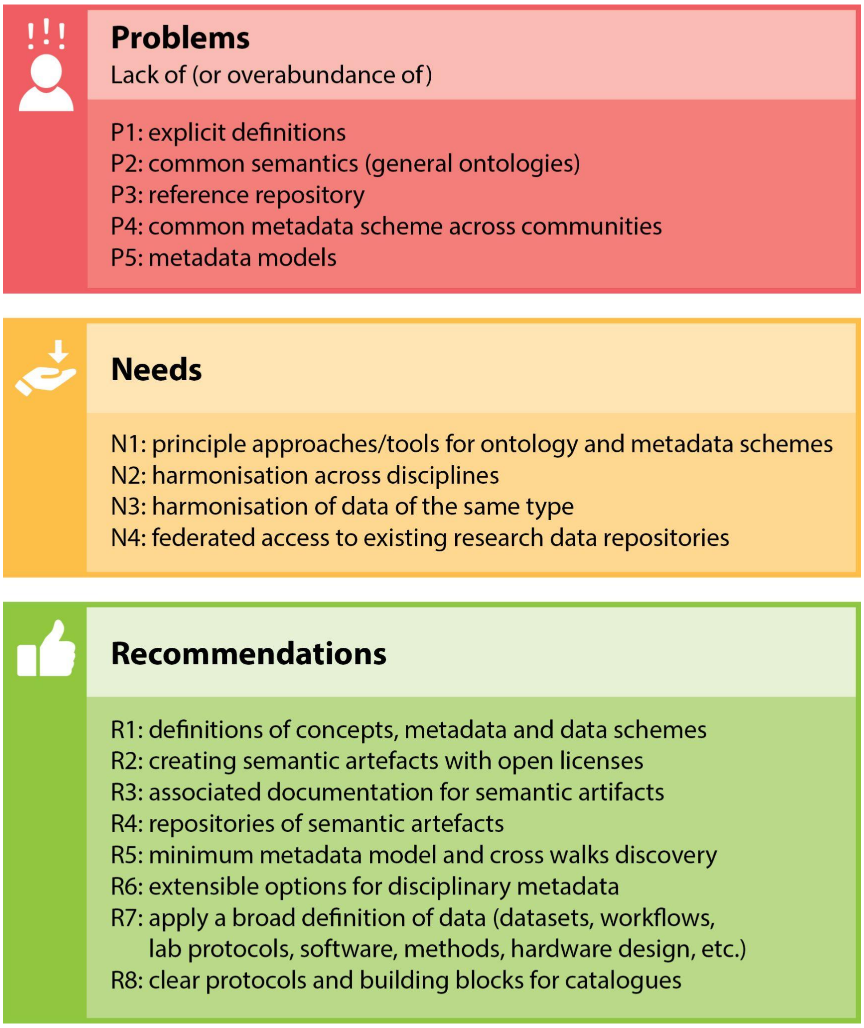

Figure 1: Schematic summary of the Problems, Needs and Recommendations concerning Semantic interoperability (ESOC-IF)

Partly, complexities also become visible in a different use of terminology. While we encourage terminological rigour, we also acknowledge that in inter- and cross-disciplinary collaboration recognising different conceptual frameworks and their related terminologies is an important part of the process. Where possible, in case of ambiguous terms we illustrate

[5] To enable cross-linking, we numberised the bullet points in the EOSC-IF section with Problems P1-P5 (e.g. lack of common explicit definitions = P1), with Needs N1-N4, with Recommendations R1-R8, and where appropriate refer to them.

our understanding by examples. Working towards a shared understanding expressed in a glossary will become one output of the TF.

(1) **What are data?** What is meant by data is by no means agreed across all academia. Data are defined in concrete research contexts, and while there is evidence that data can be repurposed, to support re-use by automated means is not trivial (Borgman 2015; Gregory 2021). Data stem from both research and industrial fields, all coming with their own information practices. The EOSC IF report departs from an even broader definition, including all varieties of Digital Objects (such as datasets, images, software, workflows). This TF acknowledges this diversity by looking into concrete practices of data (re)use and data service provision; and analyses how and to which extent Semantic Interoperability is operationalised in them. (P1, P4)

(2) **How are data described** (or the eternal question of metadata)**?** The distinction between metadata and data is contextual and depends on the research question one is following. However, in practice one differentiates between the 'data' itself, and descriptions <u>about</u> the data. For the latter, Knowledge Organisation Systems[6] are used to order information. More recently, machine executable KOS have been also labeled semantic artefacts (Coen 2019). While there exist standards (e.g., in the bibliographic domain[7]), re-using them, adapting them and integrating them into a wider framework is a research task in itself. For Semantic Interoperability (which may involve the construction of knowledge graphs) the challenge is to find a balance between genericity (overall discoverability and re-use, common semantic artefacts across domains (P2), generic metadata frameworks p. 33 EOSC-IF) and preserving the precision of tailored/targeted highly specific descriptions (high-granularity metadata). There is also a tension between granularity of the description and the contextual needs of data consumers. Providing fine-grained data is expensive and not always necessary, but meeting the contextual needs of data consumers usually results in fine-grained (meta)data descriptions.
The TF addresses this complexity by starting from the recommendations of the EOSC-IF report (p. 27) on how to best ensure semantic interoperability and to follow how those have been implemented by means of concrete case studies. Ultimately, the group will work towards a new set of recommendations for Semantic Interoperability. This work also represents a pathfinder through existing best practices and guidelines, while flagging out areas of missing or conflicting definitions and approaches. Ultimately, it is aimed at contributing to a second iteration of the EOSC - IF report.

(3) **Which new technological ways to describe and process data emerge?** While (1) and (2) encompass processes to bring data and their description to the existing technologies (adaptation), at the same time, those technologies are also changing. The TF addresses this by creating synergies between Technology Exploration Studies

---

[6] The term "Knowledge Organisation System" is used here as an umbrella term for of systems helping to order knowledge from folksonomies, thesauri, classifications, to ontologies. (Smiraglia 2016) This way we link to Knowledge Organisation as a field in information science (see ISKO.org)

[7] "A **bibliographic record** is an entry in a bibliographic index (or a library catalog) which represents and describes a specific resource" (Wikipedia). Bibliographic domain refers to libraries and archives, independently if they are hosted in academia, industry or society at large.

(PoC, case studies), which are ongoing in current and future EOSC-related projects. Next to creating cross-links among experts, the TF will also ensure that lessons learned are summarised and reflected on for the more general level of the recommendations.

(4) **How to best address Semantic interoperability as a practice? -** To achieve Semantic Interoperability is not a 'mere' process of defining suitable standards among experts and to get communities on board to implement them. Many communities are involved to explore, define and implement standards: computer and information scientists, information professionals, and diverse research communities. The latter oblige to different epistemic norms and values, which shape the way data are perceived, documented, described, and made FAIR. Moreover, institutions concerned with data (from repositories to ESFRIs) operate under different governance settings. The TF relies on methods from science and technology studies to reflect about the process around Semantic Interoperability, with the aim to support communication across knowledge domains in the making of standards for Semantic Interoperability and its implementation (training, consultancy). It is this kind of reflexive level on which we also aim to reach out and connect to those TFs which look into social, political and economic boundary conditions in bringing the EOSC to life.

Main aims

Data need to be described in a way that they can become increasingly woven into a web of data from different sources, machine operable (EOSC IF 2021). In addition, data need to be **interoperated** with tools/applications or workflows for analysis, storage, and processing. Therefore, we say (1) (meta)data use a **formal, accessible, shared, and broadly applicable language** for knowledge representation, (2) (meta)data use **vocabularies** that follow FAIR principles, (3) (meta)data include **qualified references** to other (meta)data, (4) (meta)data are at the intersection of semantic, syntactic, technical, and pragmatic interoperability.

Objectives

O1: Explorations into Semantic Interoperability

O1.1: Select/Recommend common metadata standards for a broad range of 'data' (e.g. expanding work of Ojsteršek. 2021)

O1.2: Suggest next steps in the development of catalogues for metadata standards

O1.3: Evolve syntactic interoperability for data formats, metadata schemas and services

O1.4: Support the alignment/matching of semantic artefacts, both at the domain level and at the top level, e.g., by recommending how crosswalks across them should be implemented and enacted

O1.5: Evolve the descriptions of the technical components for semantic interoperability that were identified in the original EOSC IF

O1.6: Propose strategies for long-term preservation of semantic artefacts

O2: Knowledge exchange around Semantic Interoperability

Develop trainings and/or organise workshops both to introduce the topic of Semantic interoperability and to foster cross-expert knowledge exchange (hackathons)

O3: Recommendations

New iteration of the EOSC IF - Contribute with a new set of recommendations for Semantic Interoperability.

Core activities of the Task Force

Scope: The TF aims to create synergy among already existing work which touches on Semantic Interoperability. More concretely, it departs from ongoing EOSC projects, and other projects members in the TF are involved. This ongoing work builds the core of **O1 Explorations**. The specific problem-solving activities are rephrased as “Case studies”.
We aim at organising workshops to exchange knowledge (on a general level) and to collaborate on very concrete tasks (hackathons) (**O2 Knowledge exchange**). The organisation of the workshops could be accompanied by bibliometrically enhanced literature exploration and surveys among stakeholders. This will serve to create an inventory of R&D activities relevant to Semantic Interoperability and viewpoints on Semantic Interoperability. Eventually, in the second half of the TF’s lifetime, a specific subgroup will be formed, dedicated to formulate the next iteration of recommendations (**O3 Recommendations)**.

Relevant stakeholders - partly also a reservoir for new TF members, are:

- Infrastructure and tool providers who are willing to implement parts of the EOSC IF (semantic interoperability components).
- e-Infrastructure providers, including those willing to describe and exploit Digital Objects and their metadata.
- ESFRI Clusters who may need guidelines on how to manage their semantic artefacts and Digital Objects.

Figure 2: Schema of the workflow of the TF: from case studies to a second iteration of EOSC IF

Planned duration of the Task Force

Figure 3: Timeline and stages of the TF

Working methodology

Explorations/Case Studies

The task force will apply a case-study explorative methodology, heavily relying on work in ongoing projects, networks, consortia etc.. In the TF, we currently represent initiatives such as European Health Data and Evidence network (EHDEN)[8], ELIXIR[9], Novel EOSC Services for Emerging Atmosphere, Underwater & Space Challenges (NEANIAS)[10], Research

[8] https://www.ehden.eu
[9] https://elixir-europe.org/about-us/what-we-do
[10] https://www.neanias.eu

Object-related Services for Earth Sciences (RELIANCE)[11], Social Sciences and Humanities Open Cloud Project (SSHOC)[12], The European Science Cluster of Astronomy & Particle Physics ERIC (ESCAPE)[13], European Materials Modelling Council (EMMC ASBL)[14]; Digital Infrastructure for the Arts and Humanities ERIC (DARIAH)[15], Europlanet 2024 Research Infrastructure project (EPN-2024-RI)[16], Ontology-driven data documentation for industry commons project (OntoCommons)[17], Semantic Interoperability in eGovernment Community (SEMIC)[18], German National Research Data Infrastructure (NFDI e.V.; therein, NFDI4Cat and NFDI4Ing)[19], Cross-Domain Interoperability of Heterogeneous Research Data (GO-FAIR Inter)[20], Fostering FAIR Data Practices In Europe project (FAIRsFAIR)[21], OpenAIRE[22], OpenCitations[23], and the Open Research Knowledge Graph (ORKG)[24]. These initiatives stand for a variety of academic domains (from physics to humanities), governmental and industry data - and generic, technology oriented, across domain networks. To bring actors from these fields together is already an achievement. The TF will remain open for participation from domains/sectors not yet covered.

The case studies are aligned to the Objectives, and are designed as concrete explorations of what has been formulated in the EOSC IF as recommendations for semantic interoperability.

Examples for problems to be addressed by means of case studies:

- Data – homogeneous entities and heterogeneous formats/languages (O1.1, O1.4)
- Data – human understanding and consumption (O2)
- Data – Data model versioning (O1.1, O1.3)
- Query – Federation with heterogeneous access points (O1.3, O1.5)
- On-the-fly ontology (term) integration – user-driven ontology (O1.4, O2)
- Guide users to add data in a machine-actionable way (O2)
- Interoperability of electronic lab notebooks with digital infrastructures (O1.3, O1.4)
- Assisted specification of data provenance metadata (O1.1)
- Combine discoverability across domains and findability in domains concerning (O1.2, O1.3)
- Develop visualisation to navigate through and communicate about the KOS universe (O1.2)
- Long-term archiving of semantic artefacts (cost-benefit analysis) (O1.6)
- Provenance - Audit trail of ownership (history) and transformation of open metadata with DP's (O1.4, O1.5)

The case studies operate on different levels of the EOSC infrastructure and can concern data providers work, efforts to create an open data hub, or lie with general or domain specific

[11] https://reliance-project.eu/
[12] https://www.sshopencloud.eu
[13] https://projectescape.eu
[14] https://emmc.eu
[15] https://www.dariah.eu
[16] https://www.europlanet-society.org/europlanet-2024-ri/ , in particular the VESPA WP.
[17] https://ontocommons.eu/about
[18] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/about
[19] https://www.nfdi.de/en-gb
[20] https://www.go-fair.org/implementation-networks/overview/go-inter/
[21] https://www.fairsfair.eu
[22] https://www.openaire.eu
[23] http://opencitations.net
[24] https://www.orkg.org/orkg/

service providers. Detailed description of these use cases can be found in the APPENDIX (Cases studies)

Methods

The methods used in case studies encompass:
- Desktop analysis of existing documentation (EOSC IF, Glossary of Semantic Interoperability in context of EOSC)
- Concrete research and development work (scientific publications, data sets)
- Community-engaging methods
    - Survey among EOSC relevant projects
    - Organising of outreach events (half a year infoshare)

Output

The output will take the form of
- Scientific articles and/or Conference presentations (relevant conferences)
- Event organisation (workshops with other TF's)
- Report and guidelines (Progress/Data-, Service-provider guidelines, SLA's)

Dependencies

This paragraph briefly describes dependencies in the main EOSC project to tasks and TF's . The list is condensed but not closed:

- **Standards and technical specs (EOSC Task 3.1.4/3.2.1)**: Coordination international standards; elaborate interoperability by "best practices" taken up by relevant EOSC stakeholders.
    - Data format, metadata schemas/catalogue, and common protocols.
- **Architecture and IF (EOSC Task 3.1.2) :** Building blocks of the EOSC-core, and roles of "users" in EOSC (e.g. Data Provider, Service Provider, Researchers (target audience)).
    - Infrastructure, semantic artefact deployment (Ontology - Mapping Schema - Storage (Triple store)).
- **Architecture and Interoperability Guidelines (EOSC Task 3.2.2):** Elaboration of the EOSC resource developing architecture and interoperability framework; for example guidelines for Data Provider (dataformat, data quality) and Service Providers (SLA's - how to use services, reselling aspects within the NREN, organisation).
- **EOSC IF - TF's:** Functional requirements from cross-community use-cases compatible with technical interoperability with the EOSC core.
- PID TF (other TF's
- **Train the community**: Guidelines (DP/SP/Users:Researchers...). Also training if we go into direction of guidelines

Membership

In this section, we briefly describe the competences and needed skills required for participation in the Task Force. The table below lists the current members and their skills. The TF aims at a broad range of competences and skills, focussing on computer and information sciences. The various lists are not closing and may be extended at any time. The TF aims to foster participation of Early Career Stage researchers.

Competencies:

- Ontology Know-How
- Data engineering
- Formalisation and Documentation, with specific emphasis on Knowledge Organisation
- Metadata standards
- Controlled vocabularies
- Standards, best practices, experience with open standards
- Software Engineering - full stack developer
- Project management skills
- Data Engineering, Data Science, ML and AI Know-How
- Infrastructure Know-How, Network Engineering, UI/X Design, Web know-how
- Science and technology studies, focus on interdisciplinarity and infrastructures

Skills needed:

- Standard/Ontology-specialist for defining semantic artefacts, experienced formalisation process etc.
- Full stack software engineer developing tools/applications
- Infrastructure architect deploying (data) architecture
- Stakeholder management looking for TF's make synergies and optimization
- Education/training/consultancy experiences to work with envisioned adopters….
- Experiences with analysis and intervention design for infrastructural work

TF Members (Current composition - open list)

Competencies	TF-Members
Ontology know-how	Oscar Corcho
Science studies / Infrastructure	Andrea Scharnhorst
Infrastructure know-how / Project Mgmt / Ontology Know-How / Data Engineering…	Kurt Baumann
Science studies / Metadata standards / Controlled Vocabularies / Distributed Infrastructures	Sadia Vancauwenbergh
Ontology know-how / Infrastructure Know-How / Web Know-How / Data	Silvio Peroni

Engineering / Experience with open standards / Software Engineering	
Standards, experience with open standards / Data Engineering / Distributed Infrastructures	Marco Molinaro
Formalization and documentation of knowledge, ontology/data engineering, research data infrastructures	Martin Horsch
Metadata standards / Controlled vocabularies / Project Management	Thomas Jouneau
Ontology know-how / experience with semantic databases and applications / data engineering	Lars Vogt

References

Coen, Gerard (2019) Introduction to Semantic Artefacts. Presentation given at the FAIRsFAIR Workshop on "FAIR Semantics" 22 Oct 2019. 10.5281/zenodo.3549375

Corcho, O., Eriksson, M., Kurowski, K., Choirat, C., Van de Sanden, M., & Coppens, F. (2021). *EOSC Interoperability Framework. Report from the EOSC Executive Board Working Group (WG) FAIR and Architecture*. Brussels. https://doi.org/10.2777/620649 (Abbreviated as **EOSC IF**)

FAIRsFAIR. (2020). *FAIR-Aware*. https://fairaware.dans.knaw.nl/

Ojsteršek. (2021). Crosswalk of most used metadata schemes and guidelines for metadata interoperability (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4420116

Smiraglia, R. (2016). *Elements of knowledge organization*. Springer International Publisher

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

APPENDIX

Case studies (examples)

Person: Silvio Peroni
Problem: Data – homogeneous entities and heterogeneous formats/languages
Domain: CS/SW
In the context of having a federated organisation of information within the EOSC ecosystem, consider the case when two infrastructures A and B handle similar kinds of data (e.g. basic metadata information about bibliographic resources such as journal articles), that are conveying similar semantics, but they expose them in different/non-compatible formats, e.g. RDF and JSON. It is necessary to find ways to allow A to correctly interpret B's data. Crosswalks can work for this, but the problem of the potentially-unlimited availability of data models that infrastructures in the EOSC can use to express their data can vary a lot, thus resulting in a plethora of possible crosswalks that should be implemented to enable semantic interoperability of data.

Person: Silvio Peroni
Problem: Data – human understanding and consumption
Domain: cross-domain
Within the context of the EOSC, we have to guarantee semantic interoperability of the information in a way that can be machine-readable. Semantic Web formats are adequate to represent structured data appropriately for machine consumption, but often are not adequate for human consumption, due to the intrinsic difficulties of potential users (e.g. researchers in different domains) in dealing with Semantic Web technologies. Independently from the complexity of the data model used, the most requested format by researchers for sharing/exchanging/processing data is the simplest one: CSV. Thus, it is essential to provide data in formats which enable any researcher, independently on their skills, to understand and manipulate them using well-known interfaces and tools (e.g. Microsoft Excel and LibreOffice), and such a format should be also accompanied by clear semantics. [As a side note: OpenCitations provides dump of its data in CSV, Scholix (JSON) and RDF. As of 24 May 2021, CSV dumps have been downloaded four times more than RDF dumps.]

Person: Silvio Peroni
Problem: Data – Data model versioning
Domain: CS/SW
The data models used to define data that are interchanged within the EOSC can evolve in time, and can make the data released using previous versions of the same data model inconsistent with new versions of it. Thus, it is crucial (a) to bind a particular dataset with the specific version of the data model used to describe such data, and (b) to have data model crosswalks that enable to treat and understand data defined with an old version of the data model as compliant with the latest version of it.

Person: Silvio Peroni
Problem: Query – Federation with heterogeneous access points
Domain: CS/SW
For having a federated semantic interoperability among data provided by different services we need to handle and understand, potentially, data expressed with several semantic formats, but also to enable services to interact by means of different tools and technologies, such as DBMS, SPARQL endpoints, REST APIs, etc. This aspect, which is related with the semantic interoperability, directly affects also the technical interoperability among services in the EOSC. Crosswalks at a more technical level must be implemented so as to facilitate the use of services via HTTP, for instance. There are already several conceptual proposals and implemented prototypes that try to address this issue, e.g. D2RQ (for accessing DB as virtual graphs, http://d2rq.org/), RAMOSE (to expose SPARQL endpoints as REST APIs, https://github.com/opencitations/ramose), SPARQL.anything (to try to query anything with SPARQL, https://github.com/SPARQL-Anything/sparql.anything).

Person: Lars Vogt
Problem: On-the-fly ontology (term) integration – user-driven ontology

extension in an application
Domain: CS/SW
In the context of the Open Research Knowledge Graph (ORKG), we are dealing with the problem that we cannot proactively provide the ontologies that are needed for semantically describing the research contents (not the bibliographic data) of scholarly publications. The reasons why we cannot do that are manyfold, including the fact that we do not have the expertise in all the different scientific domains that the ORKG covers and that research continuously evolves and with it also its terminology demands. What we would need for the ORKG are automatic techniques/tools that would allow our users (ORKG is a crowd-sources platform that follows an approach for data input that is comparable to Wikipedia) to “import” the terms they need for their description from external resources,
with an immediate response whether this term can be integrated with the ORKG. Currently,
as far as we know, this is not possible.

Person: Lars Vogt
Problem: Machines require more contextual information than human readers, resulting in overly complex graphs that cannot be easily understood by human readers - guide users to add data in a machine-actionable way
Domain: CS/SW
In the context of the Open Research Knowledge Graph (ORKG), we are struggling with the problem of how to guide users to add data in a machine-actionable way. Most users lack the required semantics background and thus add data with only human-readability in mind. We try to counter that by using a templating system (comparable to SHACL shapes), but due to the required flexibility for data input, we are facing limitations of what is currently possible. Also, many ontologies do not come with the information we would need to make template creation automatic. And even if they provide formal definitions of terms in the form of axioms, these axioms are often too general and lack context-dependent rules/shapes. Moreover, in many cases we have to use terms from different ontologies to semantically model a statement (think e.g. of a quality being measured with a confidence interval and a measurement unit). Since such data models use terms from several ontologies, they cannot be covered by a single ontology and are thus not readily provided by them. Bottom line: the main problem is how to reduce the graph’s complexity to only information that is relevant for a human reader and visualize only this information in the UI, hiding all the additional complexity that is required for semantic interoperability. Solutions involving shapes and graph
models must be made FAIR and openly available.

Person: Martin Thomas Horsch
Problem: Interoperability of electronic lab notebooks (ELNs) with digital infrastructures
Domain: Chemistry and chemical engineering
In the chemical sciences, a great variety of LIMS (laboratory information management systems) and ELN systems are employed to support researchers at digitizing and digitalising their experimental procedures and data. The ELNs often either have their own metadata standards, in some cases they permit researchers to create or import specific semantic artefacts for that purpose (in that case, however, there will be semantic heterogeneity within a single ELN). Research data infrastructures including NFDI4Cat, NFDI4Chem, and NFDI4Ing need to ingest data that are processed by ELNs. Beside semantic interoperability aspects, this concerns technical interoperability by agreed APIs and protocols as well as connecting multiple interoperability aspects, including pragmatic interoperability (i.e., minimum standards in practically dealing with data).

Person: Martin Thomas Horsch
Problem: Assisted specification of data provenance (cognitive process) metadata
Domain: Computational sciences and engineering
Provenance metadata are a basic requirement for making research data FAIR. They describe the cognitive processes (research workflows) by which research data were generated and thereby provide a foundation for trusting the research outcome (epistemic grounding) as well as all major aspects of reusability. A variety of semantic artefacts for provenance metadata exist; in materials science and engineering, this includes MODA for materials modelling, CHADA for materials characterization, and a mid-level ontology (PIMS-II) of the European Materials and Modelling Ontology (EMMO), all of which are mentioned in the present Horizon Europe work programme. The Review of Materials Modelling (RoMM), a compendium created within Horizon 2020, contains an annex of many workflows (from H2020 projects) annotated in MODA, which were created by hand and are somewhat idiosyncratic. Also, a MODA editor exists in an early version, and OSMO was created as an ontology version of MODA. The main challenge concerns making such metadata standards practically viable by assisting

researchers at annotating their data, since research workflows (particularly in modelling and simulation) can become very complex.

Person: Andrea Scharnhorst
Problem: Combine discoverability across domains and findability in domains concerning artefacts in long-term archives
Domain: CS/SW - application areas Digital Humanities
One increasingly popular Trusted Digital Repository Platform is *Dataverse*. In the SSHOC project, the implementation of a European Dataverse is envisioned as one of the designated repositories. Dataverse as a platform is driven by an open community approach, runs on API access, and enables dockerization of the application. It is not LD based, but very versatile concerning extensions of its Core Metadataset. This includes connections to controlled vocabularies hosted by KOS service providers elsewhere. There are two problems which could be addressed in this case: - implementation of a federated search; and - definition of a core metadata set for Dataverse instances operating in the humanities. The first problem concerns finding a shared core set of metadata; the second to navigate between such a core set and the unlimited necessities to deep-index humanities data.
EOSC IF Recommendation: Minimum metadata model and cross walks for discovery

Person: Andrea Scharnhorst
Problem: Develop visualisation to navigate through and communicate about the KOS universe
Domain: information science
Semantic interoperability encourages, not to say requires the re-use of standards, vocabularies. At the same time, there is a need in providing stable and maintained registries, catalogues for KOS, and in particular vocabulary used for Linked Open Data (see LOV https://lov.linkeddata.es/dataset/lov/). Different other tools/platforms have been experimented with examples are: https://hive2.cci.drexel.edu and https://skosmos.org. Still, for a newcomer, it is not easy to grasp the size, structure, age, liveliness (versions and maintenance), and internal structure of the Controlled Vocabulary. Standards have been the subject of visualisation research (e.g., Places and Spaces exhibit, https://scimaps.org/mapdetail/seeing_standards_a_v_130) but there is not much research about using visuals to gain an overview about a certain vocabulary. Most vocabularies are networks of information, and here a link to current research on visualising multi-level graphs might be of use. Candidates for visualisation could be taken from the KOSobservatory experiment (https://knoworg.org/the-dans-koso-observatory/)

Person: Andrea Scharnhorst
Problem: Long-term archiving of semantic artefacts
Domain: CS/SW
Semantic artefacts as services for a networked infrastructures are provided in different ways encompassing commercial and public KOS service providers but also research projects and groups. For the sustainability of any networked infrastructural solution it is important to regulate the KOS service provision (as living data). But, there might also be interest to preserve semantic artefacts just as a testimony for certain epistemological achievements (as data dumbs). Semantic artefacts can be compared to languages, and to some of them we might consider as 'endangered languages' we want to preserve. Which one, preserved how, and by whom are questions not yet sorted out.

Person: Kurt Baumann
Problem: Provenance - Audit trail of ownership (history) and transformation of open metadata from Data Providers
Domain: CS/SW application area Digital Humanities/Funding Bodies
The aim is to have the most generic implementation of Provenance (e.g. linking with Prov-O) as well as good enough needed entities and types for collecting metadata (describing data to data sets) from distributed Data repositories (Organization/Sub-Organization level). So what we are looking for is finally an - **Auditability(audit trail) of ownership (history)** and transformation (identification, validation and modification) of publicly available metadata with Data Providers (DP) based on "common" data models, standards (best practices°)