FAIR-IMPACT project response to "DRAFT: Developing and implementing the semantic interoperability recommendations of the EOSC Interoperability Framework"

Authors:

Clement Jonquet (INRAE)[1]
Esteban Gonzalez (UPM)[2]
Sophie Aubin (INRAE)[1]
Nina Grau (INRAE)[1]

Commenters:

Neil Chue Hong (UEDIN)[3], Nick Juty (UNINAM)[4], Sophie Aubin (INRAE)[1], Slava Tykhonov (DANS)[5], Nina Grau (INRAE)[1], Clement Jonquet (INRAE)[1], Esteban Gonzalez Guardia (UPM)[2], Jorik van Kemenade (SURF)[6], Lassi Lager (CSC)[7], Daniel Garijo-Verdejo (UPM), Maria Poveda-Villalon (UPM)[2], Fabrice Jouanot (CNRS)[8], Nacira Abbas (CNRS)[8], Wim Hugo (DANS)[5]

---

[1] French National Institute for Agriculture, Food and Environment (INRAE)
[2] Universidad Politécnica de Madrid (UPM)
[3] University of Edinburgh (UEDIN)
[4] University of Manchester (UNIMAN)
[5] Data Archiving and Networked Services (DANS)
[6] SURF BV (SURF)
[7] Finnish IT Center for Science (CSC)
[8] French National Center for Scientific Research (CNRS)

Background and introduction

In January 2024, the EOSC Association's Task Force on Semantic Interoperability published a draft report called "Developing and implementing the semantic interoperability recommendations of the EOSC Interoperability Framework". This report aims to enrich and refine some concepts covered by the semantic interoperability aspects within the EOSC Interoperability Framework (EOSC-IF) published in 2021.

FAIR-IMPACT aims to enable FAIR by supporting the uptake of tools, approaches and methods that are coming from various sources. This support is organised across scientific communities and at multiple levels (institutional, national and European). As the FAIR principles and cross-disciplinary efforts heavily rely on semantic interoperability, FAIR-IMPACT is keen to discuss the expansion and wording of related concepts through this open EOSC community discussion.

This report describes the views and thoughts coming from FAIR-IMPACT work packages (WP) members dealing with semantic artefacts and interoperability, which correspond respectively to Metadata and Ontologies (WP4) and to Interoperability (WP6). WP4 is actively involved in scrutinising the use of semantic artefacts from different aspects: their governance, their design, their catalogues, their use for research software, their mappings and their uses along with data repositories. WP6 is striving to design and promote interoperability mechanisms across domains and institutions and foster global alignment of FAIR frameworks. By working in concert both WPs lead to better semantic interoperability with and between disciplines, and foster semantic artefact harmonisation across the EOSC ecosystem.

The FAIR-IMPACT response is organised in five main sections according to the five stated recommendations by the EOSC association Task Force on Semantic Interoperability report, which are:

1. Align emerging adaptations and implementations to the semantic view of the EOSC-IF reference architecture.

2. Identify and consolidate different approaches to representing and exchanging (meta)data with the FDO model described in the EOSC-IF.

3. Extend the EOSC-IF to include a research process perspective that can support convergence on solutions for common use cases.

4. Extend the set of semantic objects to include artefacts such as mappings and crosswalks.

5. Recognise the semantic artefact catalogue as a critical part of the long-term viability of any research data infrastructure.

Our methodology was: WP4 and WP6 members have been invited to express their thoughts on the suggested recommendations according to their expertise and experience in FAIR-IMPACT. Thus, WP6 members mostly reviewed the interoperability aspects, specifically the first three recommendations while WP4 members mostly focused their efforts on the last two recommendations in semantic objects. But this is not exclusive. From the WP members individual points of view, feedback was gathered and synthesised, recommendation by recommendation, by one author. The overall document was then reviewed again by WP members and the FAIR-IMPACT executive team.

FAIR-IMPACT personnel who have been explicit members of the Task Force did not participate in the response.. Namely: Yann Le Franc (e-SDF), Oscar Corcho (UPM), Carole Goble (UNIMAN), Andrea

Scharnhorst (DANS) and Hilde Orten (SIKT). There could have been individual exchanges from members through other channels, but they are not publicised here.

FAIR-IMPACT's response

General Comments

As a general note, FAIR-IMPACT's comments and feedback to the TF's Recommendations are overall positive and encouraging. We have the feeling the two bodies are working towards the same direction and share the same vision of an enhanced semantic interoperability for EOSC. We have however observed that the *suggested actions* tend to be too generic, which may be subject to interpretation, reducing their effects. Our recommendation is to add concrete achievable actions that may be illustrated with examples too. Also, maybe differentiate between actions that can be executed in the short term, and those that can be implemented in the long term and will require other research and development activities within EOSC communities. This remark can be applied to the *indicators of success* too. We have identified the necessity of concrete and measurable indicators. Also, we have not necessarily identified a clear linkage between actions and indicators within a Recommendation and inter-Recommendations. For example, FAIR-IMPACT likes the focus on mappings and semantic artefacts catalogues between Recommendations #4 and #5.

Some commenters expressed the lack of actions aimed at enhancing literacy and proficiency in semantics and the use of semantic artefacts. It is crucial to grasp the significance of semantics, understand how to develop or reuse such FAIR semantic artefacts, and integrate them into information systems. Additionally, we might direct the Task Force's attention to a pertinent report[9] from the RDA Agrisemantics RDA Working Group, which offers concrete actions and recommendations for policymakers and funders (typically EOSC) pertaining to semantics.

1. *Align emerging adaptations and implementations of the EOSC - IF reference architecture*

We had feedback from 6 people on this topic. The commenters expressed that the Recommendation is somehow a call for commitment to the EOSC-IF, which is very positive. The commenters broadly agree that the main weakness of this Recommendation is the vagueness of the *indicators* and *actions*. The discourse is very general and offers few concrete elements that could guide the implementation of the EOSC-IF framework. The authors of the report try to align different emerging and divergent implementations instead of providing strong guidance both on a technical and executive level (best practices). The EOSC-IF framework is already a solid theoretical foundation; concrete pointers need to be articulated. We recognize here a limitation in the EOSC-IF: the fact that it stays very theoretical and does not rely or build on some clearly identified –and available– technologies. In other words, Recommendation #1 somehow suggests letting multiple possible implementations of the IF "emerge" and work to align them. If those implementations/adaptations are emerging, maybe a stronger requirement would have been to rely from scratch on some technical commitments rather than watching the diverging implementations emerge and then work on their alignment..

The commenters also expressed some concerns about some indicators:

---

[9]

- The indicator "Increased awareness," is extremely hard to quantify. How do we plan to measure awareness within the EOSC community? In this case, we can focus on EOSC resources implementing the EOSC-IF.

- In the case of the indicator "Shared framework for alignment", the EOSC-IF already serves this purpose. Guidelines are published in the marketplace (https://search.marketplace.eosc-portal.eu/search/guideline?q=*) and can be used by anyone to interact with EOSC services.

- Most of the commenters agree that the expression "shared vocabulary" was a bit misleading as it supposes there will be "one" shared vocabulary and suggested using the plural form.

Finally, there are other initiatives such as the Cross-domain Interoperability Framework (CDIF) that are working on similar problems. As an action, it would be interesting to analyse this framework to incorporate some elements of it in the EOSC-IF.

2. Identify and consolidate different approaches to representing and exchanging (meta)data with the FDO model described in the EOSC - IF

We had feedback from 7 people on this topic. Our main comments are related to the FDO approach. There is not a clear real implementation of the FDO Framework[10] , which makes it difficult to define clear indicators and actions. Attaching the destiny of "interoperability and FAIR" within EOSC to the destiny of the FDO Framework appears risky and may jeopardise addressing these key aspects within EOSC. Some FAIR-IMPACT members commented and discussed that the technologies to build a "Web of FAIR data and services" are certainly already available especially considering the Web and Semantic Web technologies standardised by the W3C.

In terms of interoperability, we can apply different models based on the nature of the digital objects involved. Data interoperability is the most frequent subject of interest, but the interoperability between data and software or between services has not been explored enough. The EOSC Future project[11] has been working on the creation of metadata to describe services[12] to onboard services in EOSC. We have examples of metadata for research software in one of the FAIR-IMPACT deliverables[13].

Commenters agree that there are indicators which are difficult to measure. For example, "increased availability and interoperability" and "complete set of semantic artefacts necessary to decode and make sense of their contents". The indicators are more readily understood as objectives to be achieved rather than as indicators, implying an ability to quantify them. The discourse naturally centres around metadata and semantic artefacts upon which metadata draw their alignment capability.

In the second and third indicators, the term "references" implies that the optimal approach involves (meta)data directly referencing semantic artefact content. The key challenge lies not in merely citing semantic artefact content (i.e., semantic annotations) but rather in encoding data or metadata with them (i.e., representing the data directly using the terms provided by the vocabularies/terminologies or adhering to the schemas defined by the ontologies. FAIR-IMPACT suggests to merge second and

---

[10] We may recommend not to use the expression "FDO model" which is not the one used by the DO developers. They talk about the "FDO Framework (https://fairdigitalobjectframework.org/ and https://www.go-fair.org/today/fair-digital-framework/).

[11] https://eoscfuture.eu/

[12] https://wiki.eoscfuture.eu/d play/PUBLIC/v3.00+EOSC+Resource+Profile

[13] https://doi.org/10.5281/zenodo.8199104

third indicators into one with the expression "qualified and resolvable references" and then maybe make another third indicator such as:" (Meta)data is increasingly encoded or represented with concepts and classes defined in semantic artefacts or rely on their schemas."

It is not clear in the action "converge on a common way" who should be responsible. And if the TF decides what the "relevant aspects of the FDO model" are, do they have sufficient mandate for providers and services across EOSC to actually adopt this common way.

The second proposed action involves "creating and executing processes to enhance and elaborate on current (meta)data." In our experience, the issue frequently is not the absence of a process; researchers often have various other reasons for not engaging in this task.

The third action seems very interesting, but we agree on the necessity of having a more specific recommendation. Editors should be part of the workflow of researchers and publishers. FAIR-IMPACT thinks that it is crucial to see the lifecycle of semantic artefacts from a complete point of view from designing/building them (in "editor services"), and the role of researchers and publishers. We need to be sure tools like Protégé, VocBench and others work closely with repositories such as OntoPortal ones or OLS, NVS etc. In complement, semantic artefact governance approaches can be found in the following report[14] created by FAIR-IMPACT.

3. Extend the EOSC - IF to include a research process perspective that can support convergence on solutions for common use cases

We had feedback from 6 people on this topic. The commenters agree on the importance of showing real use cases to see the value that may be added by EOSC services and investing in interoperability. One of the commenters' concerns, shared by others, is the use of the word "effectively" in the objective pointing out the difficulty to methodologically measure this effectiveness. Regarding indicators, the comment about measurable goals is repeated and shared among commenters.

We appreciate the specificity provided in suggested actions 2 and 3 regarding which communities can be engaged and what activities can be undertaken. While more detail is always beneficial, considering the scope of the recommendations, this level of specificity is acceptable. However, concerning suggested actions 1 and 4, we seek clarification on their relationship. Both appear to represent value-added implementations of the EOSC-IF and additional information sources aimed at addressing the "interoperability" challenge.

The last indicator of success, "new opportunities for collaboration and engagement", does not appear to be a clear measure of success. Opportunities can be quite vague to measure. In addition, we think the text could acknowledge here both FAIRsFAIR and now FAIR-IMPACT as this second project, in a sense, took over and builds on the first project outcomes, especially on the question of semantics. FAIR-IMPACT's WP4 is entirely focused on ontologies and metadata and has explicit tasks on multiple aspects of semantic artefacts (their governance, their FAIR design, their catalogues, their mappings, their use in data repositories). We suggest adding a reference to the project, exploring the resources that will be shared with the community.

Finally, we agree, the use of tools such as the Semantic Interoperability Profile (SIP), cited in the use cases section, benefits the identification of interoperability practices, favouring the alignments of implementations.

---

[14] https://doi.org/10.5281/zenodo.10287011

4. Extend the set of semantic objects described in the EOSC-IF to include artefacts such as mappings and crosswalks

We had feedback from 6 people on this topic. The commenters broadly agree on this Recommendation because of the importance of mappings and crosswalks in EOSC. Given that EOSC-IF only recommends maintaining crosswalks but does not provide any details on how to build, maintain and share them, the contribution of the present report is significant. Overall, the recommendation aligns very well with what is discussed in FAIR-IMPACT Task 4.4, which currently plays the role of a forum (with multiple public workshops and the creation of a new RDA working group) to discuss this matter.

One point of disagreement is that the report seems to state that there will be one and only one mapping repository for EOSC (the singular is used for mapping repository), which is not necessarily realistic nor desirable. We also think that the recommendations could go a bit further, which is explained hereafter.

The report has two major flaws that could limit its adoption, lead to different understandings by future adopters or simply to no concrete action. First, some suggested actions are a bit vague or too broad. A crucial factor in the adoption of better practices is the availability of intuitive tools to build and manage mappings: the recommendation does not precisely suggest any actions towards this. Also, the report has a full page on "Mapping repository" (page 16) which does talk about mappings and crosswalks, but not really about the fact of storing/hosting and serving these mappings in repositories or services. Maybe consider including in the report more elements, including suggested actions, related to the fact that EOSC needs to provide platforms and services to host, share and serve mappings (and maybe also ease their creation and curation). In the same line, we suggest that the task force explicitly recognizes and recommends SSSOM (almost two pages are dedicated to it in the TF report but not in the recommendations).

Action 2 (ensure the adherence to FAIR) could better target priority areas of action to make mappings and crosswalks FAIR, e.g., persistent identifiers or provenance information. Also we do not understand "interoperability enabler semantic artefact". In addition, action 2 introduces some ambiguity as it is not clear if you recommend storing mappings and crosswalks in mapping repositories or in semantic artefact catalogues.

Action 3 "Develop and support best practices for implementing and sharing toolsets that enable semantic interoperability.": This feels like it extends outside the scope of this Recommendation. Should it be something like: " Extend best practices for sharing toolsets that enable semantic interoperability, including entity mappings, schema crosswalks, and other essential semantic artefacts."

The other weakness of the report is that some major notions remain unclear despite the effort of the authors to propose definitions and contextual information: "objects" and "artefacts" are very abstract. So the objective or Recommendation #4 itself is ambiguous: "reusing the related semantic objects to support mediation across semantic artefacts". What are the related semantic objects? Are the objects the content (classes, concepts, properties, instances) of the semantic artefacts?The term "component" is used all along the report without a clear definition of what it is in this context, which readers without a solid EOSC literacy may need. Is "component" the same as what is called "mapping repository" in FAIR-IMPACT or the "mapping commons" in the SSSOM community? Last but not least, despite the definitions and explanations, the difference, if any, between mappings and crosswalks is not clear. Are they different? They map different types of objects? Or because they are handled and used differently? Or because (as it has been adopted in FAIRCORE4EOSC, a crosswalk is a set of mappings. This introduces some confusion in the recommendations.

Regarding the indicators, they seem relevant though it is difficult to foresee how they will be measured. Several discussed the phrase "compelling demonstrators", saying that compelling demonstrators are only compelling for their particular use case at the moment. It would be useful to collect use cases and have ways for users to look for commonalities, or for what they need, to follow equivalent or analogous strategies. At least one action could be suggested to improve the availability and findability of use cases or demonstrators.

The comments also suggested a couple of future directions. Some believe it would be beneficial to find a method to reduce the duplication of crosswalks, perhaps by mechanism to identify highly similar crosswalks and recommend to users depositing a new one to check if an existing one is suitable. Additionally, we could consider a system to version existing crosswalks if a new one closely resembles them. Also, some mappings and crosswalks in the near future can be predicted and kept up-to-date automatically.

5. Recognise the semantic artefact catalogue as a critical part of the long-term viability of any research data infrastructure

We had feedback from 6 people on this topic. Overall, the commenters mostly agree with the Recommendation and suggest a few terminology or expressions changes (see minor comments after). FAIR-IMPACT agrees that indeed the role of semantic artefact catalogues has been identified by the ESOC-IF but we do not see many EOSC projects that are working to design such catalogues as a core components of EOSC & research data infrastructures to ease semantic interoperability. Various perspectives were offered by FAIR-IMPACT members regarding the enhancement and implementation of semantic artefact catalogues. Comments agreed with outlined objectives and indicators, emphasising the importance of domain specificity and explicit governance for semantic artefact catalogues. They suggest refining indicators, acknowledging the multiplicity of catalogues, and advocating for their trustworthiness akin to data repositories. Indeed, a commenter inquired about the distinction/convergence between the proposed maturity model and existing requirements such as the CoreTrustSeal, highlighting the need for clarity on the model's specificity and potential impact, particularly in terms of self-assessment and certification. Trust is emphasised as a crucial factor in the adoption of semantic artefact catalogues, prompting calls for stronger connections to existing maturity and certification initiatives.

Suggestions were also made regarding the improvement of registration and maintenance of semantic artefacts processes, advocating for clear governance and user-friendly interfaces. Finally, the idea of selecting a common metadata standard for semantic artefacts is appealing and we think the point can be detailed by illustrating possible positive effects e.g., federation of catalogues to support search or more services. Improving semantic artefact catalogues interoperability is also noted as essential, advocating for cross-disciplinary catalogues and mapping capabilities. However, we express concern about the feasibility of covering diverse community needs with a single standard and propose allowing for community profiles to accommodate this variability.

The role of semantic artefact catalogues for long term archiving of semantic artefacts has also been commented on. FAIR-IMPACT also aimed to highlight the project's efforts in Task 4.1 concerning semantic artefact governance and the potential contribution of their catalogues to these governance models.

We find an ambiguity in the Recommendation to develop strategies for improvement, particularly regarding the notion of catalogue dependability in facilitating cross-domain interoperability. One commenter argued that interoperability should rely on the vocabularies or semantic artefacts themselves rather than solely on the dependability of the catalogues listing them.

Some minor comments expressed confusion about "strategies to address found" and sought clarification of the term 'dependable'. Additionally, we question the term 'established' regarding catalogues and advocate for clarity on funding and strategic improvements. Similarly to the expression "the mapping repository" in Recommendation #4, we suggest to use plural form in Recommendation #5 ("semantic artefact catalogues") to avoid making think there will be one and only one catalogues and emphase/recognize the importance of multiple semantic artefact catalogues addressing multiple community needs.

Finally, we suggest a parallel or connections between Recommendation #4 and #5, as some believe that designing a semantic artefact catalogue without a mapping service or vice versa is impractical. Our experience with semantic artefact catalogues in T4.2 suggests technology providers to develop integrated solutions offering both ontology and mapping repositories within the same platform. This aligns with the approach taken by, for instance, OntoPortal and NVS.