FAIR Ecosystem: key services, semantic technologies, trust and certification of services - Pillar 3

Session date: 6th of May 2020
Chair: Mustapha Mokrane (FAIRsFAIR)
Rapporteur: Bregt Saenen (FAIRsFAIR)

Information on the participants, the projects and working groups that they represent, and the spreadsheet used during the workshop can be found in the workshop report: https://doi.org/10.5281/zenodo.3953979

All recommendations and the action plan can be found on pp. 59-75 in *Turning FAIR into Reality*: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

This session is about recommendations 7, 8, 9 (priority) and 22, 23, 24.

Table of contents

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Rec. 7: support semantic technologies

7.1 In place and planned

*What have the projects **already done, developing or planning to do** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

- **EOSC FAIR WG**: semantics included in wider context of FAIR interoperability framework. The EOSC FAIR WG Interoperability Task Force, with collaboration of EOSC Architecture will publish in May an interim report *EOSC Interoperability framework* for public consultation. Semantic technologies are very important for data exchange because it includes some knowledge of the meaning of the data content, system structure and operation, usage constraints, and the underlying assumptions.
- **EOSC-Nordic**: promotion of standards. Nordic hopes to get a sense of what is useful to do from this synchronisation. Community standards, vocabularies, etc. This is a general ambition, not a specific deliverable.
- **NI4OS-Europe**: covering the long tail of science. Not a specific task dealing with semantics, but will try to use semantics and schemas that exist and in collaboration with other projects.
- **ExPaNDS**: won't be ready until Feb 2021, so no URL yet. The project is developing ontologies for the main application domains of Photon and Neutron science, with the aim being to standardise the metadata used in facility metadata catalogues. These catalogues will then be federated through EOSC. A key point is that the use of standard ontologies will ensure that the federated EOSC metadata catalogues are based not only on common syntax, but also on common semantics. The ExPaNDS ontology (i.e. D3.2) will be available as a service via EOSC. Making sure catalogues are using the same semantics. Working with PanOSC.
- **ENVRI-FAIR**: activities related to semantics in the different WPs. Training session in February on semantics and ontologies. This is ongoing in different parts of our work.
- **ESCAPE**: ESCAPE is supporting the use and development of International Virtual Observatory Alliance (IVOA) Semantics standards as part of general support of IVOA standards for Astronomy/Astroparticle ESFRIs. The IVOA has a Semantics Working Group that has produced 4 standards that are used for interoperability of astronomy data and services. (see: http://ivoa.net/documents/). This includes content descriptors and a framework for vocabularies, as well as a standard on units that is not limited to astronomy. These Semantics standards are implemented at some level in all services in the IVOA registry (and now available also via EUDAT B2FIND as a first step to inclusion of the operational VO framework into EOSC). There will be events on this, with a broader focus. We will talk to data providers in the community in this regard.
- **PANOSC**: The project is working together with ExPaNDS in this area.
- **SSHOC**: “interoperability hub” exists with a narrow view of interoperability.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- **FAIR Champion 1**: views on this recommendation as repository manager and someone involved in building vocabularies. Emphasis in recommendations that vocabularies should be supporting linked Open Data technology. Should be built on SKOS. Funding is an issue. Also, new vocabularies are not always sustainably maintained longer term. This issue should be emphasised in the recommendations, i.e. making them sustainable.
- **FAIRsFAIR**: in task 2.2 we work with recommendations and good practice for creating (born) FAIR semantic artefacts. This work is done together also with RDA VSSIG (https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html). If you are interested to discuss a minimal metadata set or best practice, please be in contact with fair_semantics@fairsfair.eu There will be workshops, etc. around this theme also by FAIRsFAIR this year. Persistent identifiers are discussed in the deliverable D2.1 and the theme will be extended in the following version of the report this summer (https://doi.org/10.5281/zenodo.3557381).

Rec. 8: facilitate automated processing

8.1 In place and planned

*What have the projects **already done, developing or planning to do** that addresses this recommendation? This should build on the information in the spreadsheet.*

- **EOSC WG FAIR**: provided an update to the information in the spreadsheet: the Interim PID document has been published since filling in the spreadsheet: https://zenodo.org/record/3780423#.XrJ1vJngqM8. The EOSC interoperability framework is based on FAIR digital objects with PID references to common machine readable semantic artefacts.
- **EOSC FAIR WG**: More recommendations may be needed on other elements to facilitate automation beyond semantics, e.g. PID (see also note in Whole Pillar-1)
- **EOSC-Nordic**: facilitating recommendation through testing of machine actionable metadata. Evaluation is done, monitoring continues throughout the project.
- **NI4OS-Europe**: evaluating maturity and working on optimisation process. Task 4.3. is focusing on this, leading to a demo deliverable D4.5 by middle of the project. Will also elaborate on publishing and harvesting repository metadata in a form that could be used to catalogue them. A categorisation of contemporary tools supporting FAIR and ORDM, applicable at the strategic, operational or technical level is provided in the deliverable D4.3 “Mapping of legal, technical and procedural tools”, http://doi.org/10.5281/zenodo.3820880
- **ExPaNDS**: Photon and Neutron facilities currently offer some data analysis services to their users. ExPaNDS WP4 is focused on integrating these existing services into EOSC, so that users can use them there. Aligning our work on this with PANOSC. A key focus is on ensuring that the data analysis services themselves are FAIR (i.e. not just the data they are analysing). For example, the data analysis services will not only be findable

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

and accessible through EOSC but also interoperable (e.g. through browser driven remote desktops and Jupyter analysis services). Reusability will also be greatly increased through standardising these data analysis services and making them available through EOSC.

- **ESCAPE**: processing of large amounts of data is so fundamental to astronomy/particle physics that it is not easy to understand the recommendation. This is addressed throughout several WPs.
- **FAIR Champion 1**: evaluating for her repository two emerging standards. ResourceSync and signposting standards to expose repository data to brokering services.

Rec. 9: certify FAIR services

9.1 In place and planned

*What have the projects **already done, developing or planning to do** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

- **FAIRsFAIR**: main focus on certification (assessment) in WP4 and WP2. Others touch upon it from different perspectives. M2.7 Assessment Report on 'FAIRness of Services' https://doi.org/10.5281/zenodo.3688762, there will be a framework published in the summer of 2021. WP4 works with CTS.
- **EOSC FAIR WG**: very interested in outcomes of FAIRsFAIR. RDA has worked on criteria for data, but results can still be problematic.
- **EOSC-Nordic**: this recommendation relates to one of two major tasks of EOSC-Nordic. Aims to support three to five repositories and currently selecting candidates.
- **NI4OS-Europe**: T4.4 deals with this. Two deliverables about a pre-production environment: first one in July 2020, other by the end of the project. Repositories to be onboarded in the NI4OS-Europe countries will have the opportunity to learn about the process and prepare for obtaining a certification if they decide to do so. Also, tools/services to be produced in WP4 already take into consideration FAIRsFAIR reports and deliverables about FAIR-aligned repositories and services.
- **ENVRI FAIR**: still in the mapping phase, but WP5 and WP7 will look into this.
- **ESCAPE**: The importance and usefulness of CoreTrustSeal certification is becoming more visible in the field of astronomy. The Centre de Données astronomiques de Strasbourg (see: http://cds.unistra.fr) has been certified and we advertise the fact that it was useful to do so - because it helps us with describing our own processes, and also because it helps us communicates and fit into the wider data sharing community.
- **ExPaNDS**: this work will happen fairly late in the project. Not started yet. Assess emerging certification schemes in terms of their relevance for photon and neutron science. Our task related to this deliverable (D2.6) will be to assess the certification schemes (e.g. CoreTrustSeal) and undertake an open self-assessment exercise of the

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

national research infrastructures against these schemes, i.e. how collected metadata and associated data management procedures contribute to certification and also what can be learned from related standards (e.g. ISO 16363, PREMIS, etc.) This task will work closely with FAIRsFAIR in relation to the objective of developing a certification scheme for evaluating FAIR data sharing and publication.

- **SSHOC**: Task 8.2 Trust & Quality Assurance will support SSH repositories in achieving CTS certification. SSHOC D8.2 Certification plan for SSHOC repositories https://doi.org/10.5281/zenodo.3725867.

Rec. 22: use information held in DMPs

22.1 In place and planned

*What have the projects **already done, developing or planning to do** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

- **ExPaNDS**: task is to make a recommendation on a common DMP framework for national research infrastructures and then develop a common DMP template for use in the facilities, aligned with similar work in PaNOSC. We aim to develop and trial an approach to active DMPs, which integrates with the experimental lifecycle and the metadata collection that happens (e.g. automatically) throughout this lifecycle.
- **FAIRsFAIR**: task 3.3 planning to use the RDA Common Standard for DMPs to expose DMP metadata and content from DMPonline to support integration use cases, e.g. making information about the DMP authors' choices of data and metadata available to the repositories and data centres they intend to use. Also intending to collect/share 'implementation stories' of such use cases, from other projects.

Rec. 23: develop components to meet research needs

23.1 In place

*What have the projects **already done** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

- **EOSC FAIR WG**: meeting research needs is an absolute overarching requirement, not limited to the items in this recommendation.

23.2 Planned

*What are the projects represented **developing or planning** to do? Again, this should build on the information in the spreadsheet: information about a planned deliverable, i.e. title, due date, short description*

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- **ExPaNDS**: D3.3 is related to the integration of national research infrastructure metadata catalogues into EOSC.
- **FAIR Champion 1**: I'm connected to EPOS and IPERION research communities to better know about their needs and data management challenges.

Rec. 24: incentivise research infrastructures to support FAIR data

24.1 In place

*What have the projects **already done** that addresses this recommendation? This should build on the information in the spreadsheet. Please check that there is a link to the concrete deliverable.*

24.2 Planned

*What are the projects represented **developing or planning** to do? Again, this should build on the information in the spreadsheet: information about a planned deliverable, i.e. title, due date, short description*

Whole-Pillar.1 What's missing in the recommendations and actions in this pillar?

What do projects do - related to implementing FAIR in the context of the EOSC - that is not covered by the original recommendations? Should it be included in an updated action plan and revised set of recommendations? Please focus on this pillar.

EOSC FAIR WG:

- Semantics is obviously one of the key components of the FAIR ecosystem so it is indispensable to have a specific recommendation about it, especially because it requires efforts and funding to define semantics and maintain it, and also because semantics interoperability is often taken as a second class citizen with regards to technical interoperability. There may be other components which could be worth additional recommendations/actions. The EOSC FAIR WG has been working with the Architecture WG on PID policies. PID services are core components of the FAIR ecosystem, and I am not sure that there is a specific recommendation or actions in TFiR on PID services. The PID Policies document could be used to think about a possible additional recommendation with actions - e.g. on the fact that one could work on how to certify PID services but there may be others.
  https://zenodo.org/record/3780423#.XrJ1vJngqM8
- The question of certification assessment by machine is a critical one since there is a need to ensure scalability, and one can think that the method is more reproducible and homogeneous than assessment by humans. There is a critical risk linked to the tool biases, especially if funding agencies or rules of participation use the results to fund or include repositories. An action about the fact that the proposed tools should/MUST be thoroughly tested, including by comparison with manual assessment, should be added, to try to mitigate the risks.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

- TFiR makes it clear that certification is not only for repositories. It would be useful to prioritize the kind of services which require certification beyond repositories (and fund in priority the huge work needed to define the criteria involving the stakeholders properly, but I think that this latter point is already in TFiR).
- Meeting research needs should be a very strong overarching requirement. It may be an issue that does not appear more prominently in the recommendations.

ExPaNDS: focus in our project not just on data itself being FAIR, but also on data services (e.g. data analysis services) being FAIR, e.g. not just the repository (Rec 9).

FAIR Champion 1: it would be good to emphasise that vocabularies should support linked Open Data technology. It is recommended to create vocabularies whose concepts are URI-based because this helps enhance interoperability with other like minded intitiatives on the web. Matching and linking flat concepts from vocabularies is less efficient and is not maximising the opportunites of the web. SKOS is one international standard that is widely used to build URI-based thesauri and taxonomies, https://www.w3.org/TR/skos-reference/ However, EOSC FAIR WG: let's be careful not to recommend specific technologies.

NI4OS: could include training required for smooth participation within the ecosystem, especially for resource and service providers. For example, NI4OS delivered the training on onboarding of services, related requirements, tools and certification on 3-4 June 2020, https://training.ni4os.eu/course/view.php?id=38 - some materials are quite widely applicable.

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558