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

Session date: 20th of May 2021
Chair: Simon Hodson
Rapporteur: Joy Davidson
Overall spreadsheet
All recommendations and action plan on pp. 59-75 in Turning FAIR into Reality

Event Host: Marialetizia Mari, Serenella Muradore (Trust-it)

Meeting attendance:

	Name	Organisation and project
1	Olivier Rouchon	EOSC Pillar
2	Maggie Hellström	ENVRI-FAIR
3	Timea Biro	DRI / NORF
4	Maria Johnsson	ENVRI-FAIR and FAIR Champion
5	Josefine Nordling	CSC, FAIRsFAIR, EOSC Nordic
6	Elli Papadopoulou	ATHENA RC / OpenAIRE and NI4OS-Europe
7	Erzsébet Tóth-Czifra	SSHOC, DARIAH-EU
8	Rūta Petrauskaitė	Vytautas Magnus University, HLAC
9	Stan Gielen	HLAC
10	Jessica Parland-von Essen	CSC, FAIRsFAIR
11	Barbara Magagna	Environment Agency Austria, ENVRI-FAIR, eLTER
12	Antica Culina	NIOO, FAIR Champion
13	Ilona von Stein	DANS, FAIRsFAIR
14	Gerard Coen	DANS, EOSC Synergy, FAIRsFAIR
15	Marjan Grootveld	DANS, FAIRsFAIR
16	Mustapha Mokrane	DANS, FAIRsFAIR
17	Kostas Repanas	European Commission
18	Hervé L'Hours	UKDA, FAIRsFAIR
19	Ingrid Dillo	DANS, FAIRsFAIR

20	Katrin Seemeyer	ENVRI-FAIR

Pillar 3: FAIR Ecosystem

Turning FAIR into Reality Pillar 3: Creating a technical ecosystem for FAIR data centers around FAIR Data Objects. It addresses semantic technologies and automated processing, as well certification of FAIR services and incentives for research infrastructures. Ecosystem components should meet research needs and information in DMPs should be used. This workshop session focused on reviewing progress against recommendations 7, 8, 9 (priority) and 22, 23, 24 (supporting actions) related to TFiR Pillar 3. Below, we outline activity against each of the priority and supporting actions in turn.

Rec. 7: support semantic technologies

"Semantic technologies are essential for interoperability and need to be developed, expanded and applied both within and across disciplines." (TFiR, page 66)

As noted in the report of the second Synchronisation Force workshop, about half of the INFRAEOSC-5b projects and ESFRI cluster projects focus their work on the support, promotion and use of standards, vocabularies and ontologies, and/or standard metadata schemes[1]. Updates provided during the third workshop showed that several projects are actively working to provide guidance and develop recommendations on data exchange standards as well as promoting existing guidelines and resources. Some projects are focusing on developing services and tools to support data exchange and FAIR data assessments. Several are carrying out assessments on the FAIRness of data holdings within repositories.

Building on earlier work, **FAIRsFAIR** has developed a set of data object assessment metrics and based on these, has provided free access to F-UJI[2] which is a tool to support the programmatic assessment of the FAIRness of data holdings. FAIRsFAIR has been working with a number of trusted digital repositories to scope and carry out assessments using the tool and refining it based on feedback.

EOSC Nordic undertook early work to identify data repositories in the Nordic region and to work to assess the FAIRness of their data holdings. The project initially used the FAIR evaluator tool but then opted to make use of FAIRsFAIR's F-UJI tool as it better met the project's needs. EOSC Nordic is carrying out quarterly re-assessments of the repositories in the sample to follow their FAIR maturity journey. Their assessment includes 7 metadata metrics and there is an emphasis on improving these areas by providing advice and practical recommendations on data standards. The project resources only allow for limited advice to be provided in relation to discipline specific semantic technologies and standards. EOSC Nordic has found that metadata related scores are not improving as quickly as hoped in reassessments to date and they are now aiming to explore some of the pain points in more detail to better understand how to support progress. While not a key focus for the project, **NI4OS** is also carrying out some assessments on the FAIRness of data holdings and recently put together a team to start implementing F-UJI assessments.

NI4OS' primary focus with regard to repositories is to provide general support and share good

[1] Ingrid Dillo, Marjan Grootveld, Simon Hodson, & Sara Pittonet Gaiarin. (2020). Second Report of the FAIRsFAIR Synchronisation Force (D5.5) (Version 1.0). https://doi.org/10.5281/zenodo.3953978
[2] FAIRsFAIR F-UJI FAIR data assessment tool https://fairsfair.eu/f-uji-automated-fair-data-assessment-tool

practices such as those as outlined in OpenAIRE's guidelines[3]. **EOSC Synergy** also has a core task to work on the technical policy level to ensure services are making use of standards. EOSC Synergy is working to make services aware of the standards and good practices outlined in the EOSC Interoperability Framework with an ultimate objective of supporting those services participating in the project to a common level. A mid-point deliverable is expected soon which will outline this work and a handbook will be shared towards the end to help those services not involved in the project. With regard to disciplinary support, **ENVRI-FAIR** is developing training for environmental researchers on how to use vocabularies as part of good research practice. A key for ENVRI-FAIR is raising awareness of semantic technologies and issues such as data provenance among the ENVRI-FAIR community which has been quite successful. A series of training events will begin in June 2021. ENVRI-FAIR will publish a report on their work in early June followed by a set of recommendations due in September. **SSHOC** has recently begun work with data communities in the Social Sciences and Humanities (SSH) to design a knowledge graph that will be discipline specific. Dataverse is being tested as a semantic gateway in SSHOC and this work will ultimately result in recommendations for implementing Dataverse at the institutional level. SSHOC is drawing on PARTHENOS project activity to develop a semantic annotation framework. PARTHENOS has an emphasis on the cultural heritage and heritage science research community. A key issue raised during the discussion on support was the need to ensure that countries and institutions that are newcomers will need additional help to get up to speed with the quickly evolving and complex landscape otherwise there is a risk that the current gap could widen.

Missing element:

- While there appears to be good progress in acceleration specialisation in some domains through the projects, there remains a need for a process for providing training and support for less advanced domains (as well as institutions and countries) so that current experience gap doesn't increase further.

A suggested follow up action from the second Synchronisation Force was that FAIRsFAIR, SSHOC, ExPaNDs and PANOSC should explore potential synergies in relation to their respective metadata catalogue activities. This was undertaken and, as a result, **FAIRsFAIR** is working closely with **PaNOSC**, **ExPaNDS** and **SSHOC** to carry out a metadata integration pilot to map domain specific standards to DCAT2 in summer 2021[4]. **EOSC Pillar** is also working with DCAT as they are developing and deploying Federated FAIR Data Space[5] which is based on the FAIR Data Point (FDP) concept. The FAIR Data Space will harvest metadata from repositories and map these to the DCAT format. EOSC Pillar is aiming to improve the Federated FAIR Data Space with semantic technologies but this is still in the development phase. Through their work on developing the ENVRI Hub[6], **ENVRI-FAIR** have been drawing on the RDA InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT)[7] work which may prove useful for other research infrastructures for improving search facet in portals.

Possible follow-up action:

[3] OpenAIRE Guidelines for Data Archives. https://guidelines.openaire.eu/en/latest/data/index.html
[4] Eva Mendez, Tony Hernandez, Angus Whyte, & Joy Davidson. (2020). D3.6 Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake.https://doi.org/10.5281/zenodo.4134787
[5] EOSC Pillar Federated FAIR Data Space. https://www.eosc-pillar.eu/news/federated-fair-data-space-space-federate-them-all
[6] ENVRI Hub https://envri.eu/envri-hub/
[7] InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT WG) https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg

- FAIRsFAIR, SSHOC, ExPaNDs and PANOSC to explore potential synergies with EOSC Pillar as work on the metadata catalogue integration pilot progresses

Rec. 8: facilitate automated processing

Automated processing should be supported and facilitated by FAIR components. This means that machines should be able to interact with each other through the system, as well as with other components of the system, at multiple levels and across disciplines." (TFiR, page 66)

As discovered during the second Synchronisation Force workshop, several projects are looking into the development of machine actionable pipelines and testing aspects of the The EOSC Interoperability Framework. In most cases, work is being carried out to support the automation of research related processes rather than research processes themselves. The exception being those communities of practice where the processing of large amounts of data is so fundamental to the research that automation of research processes are already implemented (e.g., astronomy).

Some of the current activities in this area focus on defining workflows. For example, **EOSC Nordic** are using the F-UJI tool to support the automated assessment of FAIR data assessment and are currently working to define how and when such assessments should be carried out within research workflows. **EOSC Synergy** is currently working to develop criteria for assessing software quality and maturity. They are developing a pipeline orchestrator and looking at tools available for carrying out the assessments against the criteria (e.g., F-UJI). Following this, the project will aim to develop pipelines for implementing the automated assessments and will support bading. The ultimate aim for EOSC Synergy's activity in this area is that more data and software will end up being accessible and reusable via the EOSC Portal and that data services made visible will be FAIR.

In other projects, the main focus is on automating interoperability rather than data assessment. For instance, **EOSC Pillar** is working to support automated processing through mappings between their Fair Data Space and repositories. **ENVRI-FAIR** has a focus on encouraging the reuse of FAIR processing workflows to ensure that researchers and other stakeholders in ENVRI FAIR can find data across the various Research Infrastructures (RIs). ENVRI-FAIR have developed a Knowledge Base which stores info about FAIRness assessments and, if some aspect is lacking, suggestions are provided on how to reuse existing processing workflows. ENVRI-FAIR is also participating in a wider context to progress FAIR digital object discussions with efforts to engage at both the RI and subdomain levels. A key area for discussion is semantics around 'types' to support machine actionability which will be crucial for enabling services to work with automated workflows. **SSHOC** has been working to progress both manual and automated curation workflows in relation to data curation workflows for the SSHOC marketplace.The main aim is to work towards realising semi-automated processes and a deliverable about this activity due to be published late 2021.

With regard to data management plans (DMPs), there is work underway in **OpenAIRE** to automate the processes relating to the creation and sharing in the Argos tool . For example, prior to deposit, repositories could be notified that they should expect a very large dataset. One of the follow-on actions suggested during the second Synchronisation Force workshop was that FAIRsFAIR, EOSC Nordic, ExPaNDs and PANOSC should explore potential synergies on machine-actionable metadata and DMPs. This is currently being progressed collaboratively through an Implementation Story on maDMPs being drafted in **FAIRsFAIR**.

Rec. 9: Certify FAIR services

"Data services must be encouraged and supported to obtain certification, as frameworks to assess FAIR services emerge. Existing community-endorsed methods to assess data services, in particular CoreTrustSeal (CTS) for trusted digital repositories, should be used as a starting point to develop assessment frameworks for FAIR services. Repositories that steward data for a substantial period of time should be encouraged and supported to achieve CTS certification." (TFiR, page 67)

As noted in the report from the second Synchronsiation Force workshop, the large majority of projects are involved in activities that will contribute to the implementation of this recommendation to varying degrees. In general, there is an emphasis on certifying repositories as service providers in the FAIR ecosystem. With regard to the certification of FAIRness and FAIR-enabling services, suggestions were made during the second workshop that metrics should be both quantitative and qualitative and that as metrics are still being defined and tested, they should be used at this stage to provide guidance on constructively improving services. Work is underway to address the suggested follow up action identified in the second Synchornisation Force workshop that there is a need to explore for what type of data services FAIR-enabling assessment and certification is valuable and feasible, in addition to repositories.

FAIRsFAIR continues to support a cohort of 10 repositories to reach CoreTrustSeal (CTS) certification. In addition, FAIRsFAR is developing its concept of CTS+FAIR which is an extension of CTS that better aligns with the FAIR principles. In addition to supporting repository certification, FAIRsFAIR has developed a draft assessment framework for assessing a border range of FAIR enabling services[8]. A workshop was held to introduce the draft framework on May 20, 2021[9] and to allow for community feedback to be collected. The draft framework is open for public comment until July 31, 2021. **EOSC Nordic** is similarly working with 10 repositories to either achieve CTS certification or to carry out self-assessments against the CTS criteria. As part of their work, one or two of those achieving CTS certified status will also test . The work is anticipated to take place in autumn 2021. **ENVRI-FAIR** is working on defining requirements and common elements among the individual RIs who are working on repository certification. ENVRI-FAIR contributing to **FAIRsFAIR** work on developing a FAIR enabling service assessment framework. **SSHOC** is currently working with 14 repositories to progress toward certification. A deliverable related to this work has been published and aims to help repository managers consider whether they should seek certification. The report aimed to focus on inclusiveness and accordingly includes information targeted to smaller repositories. Further outputs related to the certification support in SSHOC are due in early 2022. The discussion focused on whether different repositories can be certified with the same set of requirements. For example, should the same set of certification

[8] Koers, Hylke, Herterich, Patricia, Hooft, Rob, Gruenpeter, Morane, & Aalto, Tero. (2020). M2.10 Report on basic framework on FAIRness of services (Version 1.0). This draft is open for public comment until 31 July 2021. https://doi.org/10.5281/zenodo.4292598

[9] FAIR-enabling Services: Validating The Framework. https://fairsfair.eu/events/fair-enabling-services-validating-framework

criteria be applied to those repositories that are operated at the national level or by ESFRIs versus those that are at run at the institutional level.

Missing element:

- There is a gap between what is expected by CoreTrustSeal and what many institutional repositories are ready to demonstrate. While SSHOC's work picks up on this a bit (e.g., how to engage smaller repositories) there remains a need for greater clarity on whether and how smaller repositories should engage with certification processes.

A few projects are working to develop digital badging schemes. **EOSC Synergy** has carried out a state of the art review regarding digital badge issuing in relation to FAIR data assessments. The EOSC Synergy module for FAIR assessment could potentially provide a digital badge based on a F-UJI assessment. FAIRsFAIR is similarly investigating the badging of datasets that have undergone F-UJI assessments with work beginning in spring 2021.

Several projects are working to provide general support in this area. For example, **OpenAIRE** is working to align their existing guidelines against FAIR to support services to assess their FAIR-enabling capacity while **GO-FAIR** is working to support the certification of vocabularies and metadata schemas.

For some projects, repository certification is not a core activity however there is interest to see an increase in the availability of certified trusted digital repositories (TDRs) to ensure that the EOSC has a larger pool of FAIR data to make accessible for reuse. For instance, **EOSC Pillar** considers certified repositories as essential components of the FAIR ecosystem for feeding into their FAIR Data Space and they continue to monitor FAIRsFAIR activity around certification.

Rec. 22: use information held in DMPs

"*DMPs hold valuable information on the data and related outputs, which should be structured in a machine-actionable way to enhance reuse. Investment should be made in DMP standards and tools that adopt common standards and support 'active' DMPs to enable information exchange across the FAIR data ecosystem." (TFiR, page 72)*

As noted in the report from the second Synchronisation Force workshop, only some projects are actively working in this area. **ExPaNDs** and **PaNOSC** continue work to develop and trial an approach to active DMPs, which integrates with the experimental lifecycle and automated metadata collection throughout the lifecycle. **EOSC Synergy** provides some process guidance for developing DMPs for services joining the EOSC. **SSHOC** does not actively focus on DMPs in their project but are trying to promote and support the publication of DMPs among the SSH domain. The participants stressed that as many communities are still in the process of implementing good practice around developing and updating DMPs there is a limit to what can be done to reuse the information held in DMPs. Until good practice is fully embedded, we will not be able to extract content from DMPs at scale.

Missing element:

- It was suggested that as community good practices are relatively new with regard to DMPs, this theme could be explored more fully in the next generation of thematic projects.

Rec. 23: develop components to meet research needs

"While there is much existing infrastructure to build on, the further development and extension of FAIR components is required. These tools and services should fulfil the needs of data producers and users, and be easy to adopt." (TFiR, page 73)

During the second Synchronisation there was a suggestion that this supporting action might be made a higher-priority in TFiR to recoginse the crucial importance of ensuring that the European Open Science Cloud is researcher-centric. This was echoed during discussions during the third workshop which focused on some of the challenges of engaging with end users communities. **ENVRI-FAIR** is working to ensure they meet researchers' needs through dedicated activity to collect researchers' requirements. The work involves collaborations between the RIs of four sub-domains and aims to come up with concrete use cases to test the FAIRness of services that are currently available as well as informing the development of those services still in the planning stage. **SSHOC** also agrees on the need to align with researchers' needs and the need to secure a better understanding of potential users' needs. SSHOC are in the early stages of user testing of the SSHOC Marketplace which will involve about 150 testers from the research communities. SSHOC also has a dedicated Data Communities work package which is doing valuable work in reaching out to current and potential users.

During the **ENVRI-Plus** project it became clear that in many cases the people involved in managing and operating the research infrastructure have one foot in the research community and one in the RI. In such cases, it is often the case that the needs of specific sub-domains represented among the RI partners are very well understood but the needs of the broader actual and potential user base of the RI are less well understood. There is also often a difference between well-versed users (esp. in large organisations) and new users when it comes to being vocal about usability needs. We must ensure that there is equally a focus on feeding in broad-based usability needs from the longer tail (smaller research groups, etc). Having one foot in research and one in the RI is something that is also encountered in SSHOC. There are also ongoing challenges associated with how we can obtain input on needs from the longer tail of the research community. It is also often difficult to secure participation in surveys and/or testing which can be time consuming for participants. **SSHOC's** forthcoming deliverable on challenges in engaging with the community is due towards the end of the project and could have valuable insights for others.

Rec. 24: incentivise research infrastructures to support FAIR data

Research facilities, in particular those of the ESFRI and national Roadmaps, should be incentivised to provide FAIR data by including it as a criterion in the initial and continuous evaluation process. Investments should be made strategically and consider data service sustainability." (TFiR, page 73)

As noted in the report of the second synchronisation Force workshop, there were few activities relating to incentivisation research infrastructures (RIs). **EOSC Nordic** is continuing to progress its mapping of existing policy incentives across Nordic countries and work to engage with stakeholders with the aim to develop and promote new and/or efficient incentives. A

number of interviews will be carried out with stakeholders over summer/autumn 2021. EOSC Nordic is also engaging with policy makers to try to harmonise the policies in this region. **GO FAIR** is working to promote the M4M rapid pipeline and handbook for research infrastructures.

Several projects are working to support RIs indirectly to be more able to support FAIR data provision such as **EOSC-Pillar's** approach to support data exchange through their FAIR Data Space which will connect data islands. Also noted in the second workshop report, incentivising research infrastructures to support the production and use of FAIR data requires active involvement and commitment from a range of national and international stakeholders. Discussions during the third workshop highlighted national roadmapping exercises as a potentially valuable tool for engaging with politicians.

The participants of the third workshop also note that while there is an emphasis placed on RIs to engage with politicians and policymakers, there may be less of an emphasis for them to reach out to their user communities to encourage them to make their research data outputs FAIR.

Missing element:

There is a need to ensure that research infrastructures are working to encourage and support their user communities not only to reuse others' FAIR data but also to commit to creating and contributing to the body of FAIR data available. Training will be crucial for realising this necessary culture change.