Incentives and metrics for FAIR data and services - Pillar 5

Session date: 12th of May 2021
Chair: Mustapha Mokrane
Rapporteur: Josefine Nordling
All recommendations and action plan on pp. 59-75 in Turning FAIR into Reality

Event Host: Sara Pittonet, Marialetizia Mari

Meeting Attendance:

	Name	Organisation and project
1	Mustapha Mokrane	DANS, FAIRsFAIR
2	Neil Chue Hong	Software Sustainability Institute / University of Edinburgh / EOSC FAIR WG / FAIR4RS WG
3	Maggie Hellström	ICOS ERIC and ENVRI-FAIR
4	Josefine Nordling	CSC, FAIRsFAIR
5	Hervé L'Hours	UK Data Service, FAIRsFAIR
6	Federica Garbuglia	EUA/FAIRsFAIR
7	Katrin Seemeyer	Forschungszentrum Juelich, ENVRI-FAIR
8	Marialuisa Lavitrano	EOSC Association / University Milano-Bicocca / BBMRI-ERIC
9	Abigail McBirnie	ExPaNDS/UKRI-STFC
10	Antica Culina	FAIRsFAIR champion, SPI-Birds
11	Antti Pursula	EUDAT CDI
12	Erzsébet Tóth-Czifra	DARIAH-EU (SSHOC)
13	Jakub Urban	CERN, ARCHIVER Project
14	Karsten Peters-von Gehlen	German Climate Computing Center (DKRZ), FAIR Champion representative
15	Hugh Shanahan	FAIRsFAIR
16	Christian Cuciniello	EC
17	Richard Dennis	UCPH
18	Marjan Grootveld	DANS, FAIRsFAIR
19	Susanna-A Sansone	FAIRsharing, ELIXIR-UK, University of Oxford; FAIRsFAIR Data Champion. Also reporting on EOSC-Life and FAIRplus.

20	Timea Biro	DRI/NORF
21	Milan Ojstersek	EOSC WG FAIR
22	Elli Papadopoulou	NI4OS
23	Iryna Kuchma	OpenAIRE
24	Sarah Jones	GEANT, EOSC Association
25	Serenella Muradore Gallas	Trust-IT Services
26	David Carr	Wellcome Trust/ HLAC
27	Barbara Magagna	Environment Agency Austria, ENVRI-FAIR, eLTER
28	Fernando Aguilar	CSIC / EOSC-Synergy

This session is about recommendations 12, 13 (priority) and 25, 26.
There are 2 questions per recommendation, followed by 2 questions about the pillar.

Rec. 12: Develop metrics for FAIR Digital Objects

12.1 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.*

12.2 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*

Neil Chue Hong (EOSC WG FAIR): In RDA/ReSA/FORCE11 FAIR4RS WG, a task force to define metrics for FAIR software is being scoped. When this is launched, it will be advertised via the FAIR4RS WG, and the RDA pages.

Marialuisa Lavitrano: **EOSC Association** has a task force on FAIR metrics and data quality, to oversee the implementation of FAIR metrics and make sure these are fit for purpose. Data should be robust enough and guided by the communities. Charters for task forces are drafted, call for joining task force will soon open, and the task forces will last for 18-24 months. It is important to add to the SRIA material, link up with relevant projects and to ensure the implementation.

Maggie Hellström (**ENVRI-FAIR**) FAIR Digital Object Forum (https://fairdo.org), back-end of FDOs, is developing a technological specification for FDOs. Metrics being developed, but they are at the most in early phases. Metrics will likely be worked on in their TSIG and BIG working groups, in the context of evaluating the readiness of a DO to be machine processed using e.g. the DOIP protocol. FAIRDO is also aware of other initiatives, e.g. F-UJI, which will be taken into account in this task force. Will be completed within about a year's time.

Mustapha Mokrane: FAIRsFAIR is working on mechanisms to extend CTS certification with FAIR elements. The relevant deliverables will become available in August 2021 and February 2022.

Abigail McBirnie (**ExPaNDS**): As part of the process of creating our draft metadata framework for Photon and Neutron research data management at PaN RIs (https://doi.org/10.5281/zenodo.4312825), we looked at what metadata is available at each stage of the (idealised) experimental lifecycle; for example, the PI name and identifier become available at the first stage of the life cycle, i.e. the proposal stage. For each metadata type, we then used the RDA FAIR Data Maturity Model prioritisation indicators to prioritise that metadata type (i.e. as essential, important, useful, not relevant) for each element of FAIR (i.e. F, A, I, R). So, this is not developing metrics in itself, but it provides an example of a (partial) reuse of an existing model.

Josefine Nordling: EOSC-Nordic did a large FAIR maturity evaluation early in the project: first used the Wilkinson tool then but moved to F-UJI, because that addressed more aspects. FAIR maturity levels in the 100 repositories in the sample haven't raised since the last assessment (every 3-4 months) so EOSC-Nordic now offers additional 1:1 guidance. EOSC-Nordic doesn't distinguish between domains in the evaluations, but will take a lead in guiding repositories in using community-endorsed metadata features.

Jakub Urban (**ARCHIVER**): Is integrating F-UJI tool to the testing pipeline in order to monitor the compliance of the repositories and uptake of the principles. ARCHIVER is also helping to adapt the tool to specifics of the research communities participating in the ARCHIVER Project and also improving the tool as they found some mitigation affecting fully evaluating our repositories.

Karsten Peters-von Gehlen: **DKRZ**: FAIR maturity assessments of their data holdings have been made using the F-UJI tool. However, they have encountered a few bugs and changes in these might affect the scores somewhat (as these will not be fully comparable).

Susanna Sansone: (reporting on **FAIRplus**): We have FAIRified a variety of life science examples: clinical trial, clinical observation, and molecular datasets. We have used the RDA/FAIRsFAIR indicators and tested the various evaluation tools. We have noticed that the tools and the indicators are too generic for our datasets and use cases, and we had to extend them and make them more specific.

Rec. 13: Develop metrics to certify FAIR services

13.1 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.*

Mustapha Mokrane: **FAIRsFAIR** published a milestone report on the FAIRness of services.

13.2 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*

Abigail McBirnie: **ExPaNDS** and **PaNOSC** work on certification of data catalogues, and an open self-certification assessment. They will be exploring synergies with F-UJI tool and CTS and other initiatives (e.g. relate to work done in PaNOSC).

Josefine Nordling: EOSC Nordic developed a maturity framework for services, but intentionally addressed FAIR hardly because of several other initiatives. FAIR aspects can be added later to the framework.

Maggie Hellström (**ENVRI-FAIR**): Please recommend in the SF report that "service" should really be defined unambiguously. We need a typology of services, as it is too confusing now, also among various stakeholders. There is the risk of misusing/misinterpreting metrics when the object is unclear, in addition to the trend that "FAIR" is being used for other resources than Digital Objects.

[A point raised very explicitly in our Second Synchronisation Force workshop, 2020] Metrics are snapshots in time. It is important that one assesses objects over time! Also important to keep a record of all the tests, so that progress can be followed in a transparent manner.

Susanna Sansone: In **ELIXIR** (life sciences), there are Core Data Resources (CDRs) and Deposition Databases, which have been selected according to a set of indicators. There are also Recommended Interoperability Resources (RIRs) (FAIRsharing is one of them), which are resources and services that facilitate FAIR-supporting activities, and also these have been selected according to a set of indicators.

Rec. 25: Implement and monitor metrics

25.1 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.*

Mustapha Mokrane (**FAIRsFAIR):** The report on 'FAIR data assessment mechanisms to develop pragmatic concepts for FAIRness evaluation at the dataset level' has been published. This includes assessments e.g. on F-UJI and FAIR Aware. A data badging schema will be developed after the FAIR data assessment part is done. Report will be published during the summer of 2021.

Jakub Urban (**ARCHIVER**): they are working on the metrics tool and open self-assessment process on CTS.

Abigail McBirnie (ExPaNDS): ExPaNDS does data policy work and monitoring around FAIR. They surveyed whether monitoring of data compliance is in place, however, this is rarely the case. Monitoring FAIR in audits and including in policies is the ideal, but we need to have the relevant practices (criteria/metrics) in place to do so.

Karsten Peters-von Gehlen (**DKRZ**): (Meta)Data submission guidelines in place at the WDCC repository do not explicitly mention FAIR, but the submission guidelines ensure a high degree of long-term reusability of data "by design". Compliance with the submission guidelines is checked by repository staff and iterated in a one-on-one communication

process with the data providers if compliance is not given at the time of submission.

25.2 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*

Barbara Magagna - **ENVRI-FAIR**: FAIR Implementation Profiles (FIP) have been implemented, which allows to compare the FAIR maturity levels, landscaping tool for looking at the FAIR-enabling resourcing used. Also the FDO Forum is to be consulted to get a broader perspective on this issue. Convergence is important and specifically interoperability, e.g. between different vocabularies. Will be developed within the course of this year.

Rec. 26: Support data citation and next generation metrics

26.1 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.*

Abigail McBirnie wrt Action 26.4: in a symposium in the autumn, **ExPaNDS** will bring PaN RI facility librarians and data managers together around FAIR and how to incentivise people to work towards FAIR. In relation to impact and metrics, one question our facilities need to consider is where we will put our focus on in the future. For example, reporting is likely to require reporting on FAIR - how will our facilities handle this?
On a more general point, it could be useful to make more active steps to bring together the bibliometrics/altmetrics/metrics community and the FAIR metrics community, perhaps through joint events or discussion. At the moment, they rarely cross paths and can be very separate communities. Yet, both address metrics in one way or another and have a lot of experience in best practice around the use of metrics.

Maggie Hellström (**ENVRI-FAIR**): More support is needed in how to cite FDOs, which ties into the licensing of data, i.e. how you define the licensing strings? Individual and organisational roles and concerns/needs have to be further defined. We need to be able to tie the relevant actors in a logical way which would make it easier to advance some remaining challenges.

Abigail McBirnie: Yes, organisations often use licensing to try and force data citation - but citation is an 'academic norm' issue while license is a copyright/legal issue. In terms of data, we have noticed a tendency for organisations to assume that what works for publications (e.g. CC-BY) also works for data. However, certainly in PaN, much of our data is measurement data so it is not even clear that copyright applies to those data - and therefore, a CC-BY is not appropriate, i.e. as it assumes copyright applies.

Erzsébet Tóth-Czifra (SSHOC): SSHOC D3.4 "Making data findable by being citable" is in the making. Citation practices are also relevant for the question how to credit intermediaries, such as a platform like Europeana, where reusers found the resources.

Josefine Nordling: EOSC Nordic will organise a FAIRification event after the summer on community-specific metadata, using F-UJI.

26.2 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.

Whole-Pillar.2 Any recommendations not addressed?

Are the recommendations being covered enough by these activities? If not: what should be done? And by whom?

Maggie Hellström: The need for transparency in assessments, temporal evaluation of assessments is important. E.g. also by explicit versioning of metrics tools with the help of time stamps. Need to stress that a lower FAIR score does not mean that the quality of data is necessarily poor and of low quality.

Karsten Peters-von Gehlen (**DKRZ**): Regarding transparency: the Wilkinson tool and also the FAIRshake tool https://fairshake.cloud make the results of any assessments available on the homepages.

Susanna Sansone: @Karsten many assessment/evaluation tools listed at https://fairassist.org

Abigail McBirnie: Reporting on FAIR is currently a non-priority recommendation. Increasingly, though, we are seeing data legislation that covers or refers to FAIR. As such, it may not be that long until we see reporting rise up as a much higher priority, i.e. to comply with legislation or the need to report at the national level (e.g. OECD reporting). Do we need to adjust the FAIR principles to take this into account, for example, by making reporting on FAIR part of a priority recommendation?

Sarah Jones: the EOSC Association started a Task Force on 'FAIR metrics and data quality'.

Maggie Hellström: Also: "complete metadata" in terms of people adding values to all metadata fields ("rich" metadata) doesn't mean that the values themselves are correct or even useful. We need to raise this point as well!