Concepts for FAIR implementation: FAIR Digital Objects and technical components of the FAIR ecosystem - Pillar 1

Session date: 11th of May 2021
Chair: Gerard Coen
Rapporteur: Bregt Saenen
Hosts: Sara Pittonet, Marialetizia Mari

All recommendations and action plan on pp. 59-75 in Turning FAIR into Reality (This session is about recommendations 1, 2, 3 (priority) and 16, 17)

Meeting Attendance:

	Name	Organisation and project
1	Gerard Coen	DANS, FAIRsFAIR
2	Bregt Saenen	European University Association, FAIRsFAIR
3	Sara Pittonet Gaiarin	Trust-IT, FAIRsFAIR
4	Marjan Grootveld	DANS, FAIRsFAIR
5	Antica Culina	NIOO-KNAW, SPI-Birds, FAIRsFAIR champion
6	Andreas O Jaunsen	EOSC-Nordic WP4, Nordforsk
7	Richard Dennis	Royal Danish Library / University of Copenhagen Library
8	Maggie Hellström	ENVRI-FAIR, ICOS ERIC
9	Katrin Seemeyer	ENVRI-FAIR, FZJ
10	Maria Johnsson	ENVRI-FAIR, ICOS ERIC, FAIR Champion
11	Ingrid Dillo	DANS, FAIRsFAIR
12	Andreas Athenodorou	The Cyprus Institute, NI4OS-Europe
13	Mari Kleemola	FSD/Tampere University, EOSC Nordic WP4
14	Arnaud Gingold	OPERAS RI / CO-OPERAS IN
15	Mustapha Mokrane	DANS and FAIRsFAIR
16	Elizabeth Newbold	STFC, FAIRsFAIR
17	Jakub Urban	CERN, ARCHIVER Project
18	Federica Garbuglia	EUA, FAIRsFAIR
19	Elli Papadopoulou	ATHENA RC / OpenAIRE and NI4OS-Europe
20	Aoife Coffey	NORF Ireland
21	Barbara Magagna	Environment Agency Austria, ENVRI-FAIR, eLTER
22	Marialetizia Mari	Trust-IT, FAIRsFAIR
23	Timea Biro	DRI / NORF Ireland
24	Karsten Peters-von Gehlen	German Climate Computing Center (DKRZ), FAIR Champion
25	Andreas Rauber	TU Wien, EOSC-Sec, FAIR Data Austria, FAIR champion

26	Christian Cuciniello	European Commission
27	Hervé L'Hours	UK Data Service, FAIRsFAIR
28	Fernando Aguilar	CSIC / EOSC-Synergy
29	Serenella Muradore Gallas	Trust-IT Services, FAIRsFAIR
30	Susanna-A Sansone	FAIRsharing, ELIXIR-UK, University of Oxford; FAIRsFAIR Data Champion
31	Abigail McBirnie	ExPaNDS / UKRI - STFC
32	Milan Ojstersek	EOSC WG FAIR
33	Mark Allen	ESCAPE, FAIR champion
34	Eoghan Ó Carragáin	NORF, Ireland

3

COLLABORATIVE NOTES:

Rec. 1: define FAIR for implementation

To make FAIR data a reality it is necessary to incorporate and emphasise concepts that are implicit in the FAIR principles, namely: data selection, long-term stewardship, assessability, legal interoperability and the timeliness of sharing.

- **Mark Allen** (champion) Science cluster Escape to map community standards in astronomy, map them to FAIR data principles. Ensure these are being implemented complementary to FAIR data standards.
- **Andreas Jaunsen** (EOSC-Nordic) rely on FAIR assessment tools, now mostly FUJI tool. Average FAIR score of about 20% among 100 repositories being looked at. Making recommendations to improve score. Identifying repositories where there has been a significant increase/jump in score after they (EOSC-Nordic) have worked with them.
- **Karsten Peters-von Gehlen** (champion; from chat) There has been a relevant update on F-UJI software fixing some bugs which I pointed at Robert Huber. FAIR scores might increase again due to that.The update took place about 2 weeks ago.
- **Andreas Rauber** (champion) only challenge for us is to push machine actionability as soon as possible. OSSDIP, the Open source Secure Data Infrastructure and services has been released in v1 (http://www.ifs.tuwien.ac.at/~andi/secure_data_infrastructure.html) offering a reference implementation of a completely isolated data visiting environment) v2, which is under development now will include a mechanism for publishing the metadata of the closed data, also including subsets selected dynamically via queries applying the RDA WGDC recommendations.
- **Abigail McBirnie** (STFC, ExPaNDS) worked with PANOSC to finalise PaNOSC model data policy last year. However, the needs of ExPaNDS facilities led to the need for a higher-level data policy framework as well. Result of this work was a draft 30 element

framework. Following data policy consultations with all of our partner facilities this year, we will present the final version of the ExPaNDS data policy framework in an upcoming deliverable in August 2021. Workshops on FAIR for facility staff in autumn 2021: FAIR for Facilities and The FAIR Experiment

- **Maria Johnsson** (ENVRI-FAIR) spreadsheet reflects ongoing work. Last year has been "practical" work.
    - **Maggie Hellström** in ENVRI-FAIR bringing together four different subdomains. Working within subdomains to map out what is necessary to increase FAIR of data. Explicitly looking at FAIR whereas before it was perhaps more subconscious activity in the sub-domains. Second step is to bring everything together in the ENVRI hub, a common platform for metadata and services. It's a bottom-up approach, but coordinated.
    - **Barbara Magagna** (ENVRI-FAIR) pre-tool for assessing FAIR profiles. Assessing approaches towards the FAIR principles. Being incorporated in ENVRI knowledge base. Third run of FAIRness assessment in 2022.
- **Jakub Urban** (ARCHIVER) setting up service for long-term preservation. Integrating FAIR principles by exploring the FUJI tool. Had several meetings with Robert Huber to investigate and evaluate the current state of our repositories. Evaluating our repositories from this perspective to guarantee FAIR principles compliance in the procured cloud services.
- **Andreas Athenodorou** (NI4OS) focussed a lot on training events and training people, as well as developing tools that would enable us to implement FAIR in a practical way. Better under the training TFiR recommendations. Also, NI4OS-Europe focuses on legal interoperability on the level of licenses for derivative works : License Clearance Tool that was developed addresses such challenges in the RDM lifecycle (https://wiki.ni4os.eu/index.php/License_Clearance_Tool_-_Description_and_Documentation)

1.1 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.*

1.2 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*

Rec. 2: implement a model for FAIR Digital Objects

Implementing FAIR requires a model for FAIR Digital Objects. These, by definition, have a PID linked to different types of essential metadata including provenance and licencing. The use of community standards and sharing of rich documentation is fundamental for interoperability and reuse of all objects.

- **Andreas Jaunsen** (EOSC-Nordic) not much to update.
- **Barbara Magagna** (ENVRI-FAIR) including model of FAIR digital objects related to nano publications. Still under development. Deliver something in the coming months.

- **Elli Papadopoulou** (NI4OS-Europe)
  *also supporting integrations that have to do with Handle and DOI where NI4OS-Europe partners have capacity and skills. Use of Argos DMP tool supports links between Argos DMP machine actionable outputs and other entities included in the OpenAIRE Research Graph, e.g. funders, repositories, datasets, metadata etc
- **Abigail McBirnie** (STFC, ExPaNDS) last year, we produced a draft metadata framework for FAIR data management in Photon and Neutron facilities. This framework considered what metadata is available at the different stages of the experimental lifecycle in Photon and Neutron facilities. We also considered the systems from where the metadata can be captured and the human actors/roles that are involved. We're trying to establish the priority/minimum metadata to make the workflow FAIR. Now will engage with facilities to see what's possible now/in future and to help produce a final version of the metadata framework. We also introduced this metadata framework idea to facility staff at the FAIR workshop series. Re PIDs, we have had a best practice sharing workshop to look at what PIDs for data are already being used in facilities. We will also be looking at advanced infrastructure for PIDs, especially around linking between PIDs for different components of the research ecosystem (e.g. people, data, publications, instruments, samples, etc.)
- **Gerard Coen** (EOSC Synergy) the documentation for the EOSC Synergy software quality assurance pipeline is available here: https://indigo-dc.github.io/jenkins-pipeline-library/release/2.1.0/index.html The pipeline validator is also available here: https://github.com/EOSC-synergy/jpl-validator
- **Maria Johnsson** (ENVRI-FAIR) practical and technical work ongoing. Regarding action 2.2 we have a lot of training ongoing throughout the project. WP6 (chaired by Maggie Hellström) is focussed on training, but this happens throughout the project.
  - **Maggie Hellström** (ENVRI-FAIR) developed a catalogue and training platform.
- **Jakub Urban** (ARCHIVER) testing the prototype services that they support Automated metadata indexing for several tens of PB content must be standard, aiming at maximum interoperability, including support for dataset filtering. Implementing Federated Identity and Access Management functionality.
- **OpenAIRE** in collaboration with La Referencia are approaching repositories enhancements with the scope of creating FAIR digital objects. Use of Argos DMP tool supports links between Argos DMP machine actionable outputs and other entities included in the OpenAIRE Research Graph, e.g. funders, repositories, datasets, metadata etc

2.1 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.*

2.2 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*

Rec. 3: develop components of a FAIR ecosystem

The realisation of FAIR data relies on, at minimum, the following essential components: policies, Data Management Plans, identifiers, standards and repositories. There need to be registries cataloguing each component of the ecosystem, and automated workflows between them.

- **Andreas Jaunsen** (EOSC-Nordic) main thing has been the execution of a series of verification events that focus, technically on various aspects of FAIR implementation. Will continue for the remainder of 2021. Upcoming workshops will focus on enriching community standards.
- **Elli Papadopoulou** (NI4OS-Europe) focus is on machine actionability. In NI4OS: a tool that attempts to create machine actionable repositories' policies has been developed and currently runs on beta version. The tool is RePoI: https://repol.ni4os.eu/.
  In OpenAIRE: Argos machine actionable tool that runs in production: https://argos.openaire.eu/splash/. Argos DMP outputs enhance the OpenAIRE Research Graph and create links with other outputs and relationships (semantics).
    - **Andreas Athenodorou** (NI4OS) summer school on Open Science to present these tools.
- **Milan Ojstersek** In the EOSC interoperability framework 2.0, we have proposed how to manage FAIR Digital Objects in the context of EOSC, and we have proposed a reference architecture for the EOSC Interoperability Framework that is inspired by and extends the European Interoperability Reference Architecture (EIRA). We have identified which main building blocks are required. We also have made an analysis of existing metadata models and an initial set of crosswalks among them. This initial work may set the initial steps for a future proposal for an EOSC Minimal Metadata Application profile, which should be widely discussed and agreed upon by disciplinary communities. EOSC interoperability framework is accessible on https://www.eoscsecretariat.eu/news-opinion/achieving-interoperability-eosc-interoperability-framework.
- **Andreas Rauber**: Focus on the integration of maDMPs with the Invenio-based repository infrastructure as part of FAIRDataAustria initiative
- **Barbara Magagna** (ENVRI-FAIR) developments of RDA working groups related to this recommendation. Focussing on developing vocabularies.
- **Jakub Urban** (ARCHIVER) Prototype demonstrators of full reproducibility of services (initial examples are database services and/or software distribution services) on top of the resulting supported data archives.
- **Abigail McBirnie** (STFC, ExPaNDS) In relation to Rec 3, we are focusing a lot on (active) data management planning (as a verb!, i.e. rather than the more static concept of the DMP). We are working closely with PANOSC in this activity. The first step is to define a template of appropriate questions. We have started with those from the RDMOinfo pool, but these need adapting for the PaN context in many cases. There are also additional questions we need to add as well as some RDMOinfopool questions that we don't need as they are not really relevant to the PaN context. In ExPaNDS, we are also doing work looking at where the information used to populate the DMP can come from as well as at why we need to collect it/what facilities are going to do with the information collected (we don't want to create an unnecessary burden by collecting

information that won't be used). In the later stages of this work, we will look at piloting active DMPs at some of our partner facilities.

- **Susanna Sansone** (standing in for EOSC-Life):
  - several FAIR components released, described in this report.
  - Also here is a EOSC-Life collection in FAIRsharing that represents more then 100 diverse data resources/repositories produced by EOSCLife partners; here is the graphical relationships between standards used by these repositories.
  - Here a crosswalk of most used metadata schemes and guidelines for metadata interoperability
- **Susanna Sansone** (as FAIRsharing):
  - Ongoing collaboration with openAIRE, DMPonline, FAIRshake, FAIR Evaluator and other harvester, DMP and assessment/evaluation tools to serve the FAIRsharing content (repositories and standards) for their consuption
  - Ongoing collaboration with FAIRsFAIR to register repositories and policies

3.1 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.*

3.2 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*

Rec. 16: apply FAIR broadly

FAIR should be applied broadly to all objects (including metadata, identifiers, software and DMPs) that are essential to the practice of research, and should inform metrics relating directly to these objects.

- **Andreas Jaunsen** (EOSC-Nordic) plans to collaborate with WP3 in EOSC-NORDIC on services.
- **Elli Papadopoulou** (NI4OS-Europe) trying to foster policy adoption, incl. FAIR, in the context of the project's objective to set up National Open Science Cloud initiatives (NOSCIs) in the 15 countries of its partners. "A National Open Science Cloud Initiative (NOSCI) is envisaged as a coalition of national organizations that have a prominent role and interest in the Europe Open Science Cloud (EOSC). The main aim of NOSCI will be the promotion of synergies at national level, and the optimization/articulation of their participation to European and global challenges in this field of Open Science, including the EOSC. National initiatives are envisaged to play a prominent role in Member States and Associated Countries and facilitate EOSC governance."
- **Abigail McBirnie** (STFC, ExPaNDS) in autumn an event for facility librarian and data managers, aimed at linking these two (often separate) areas in our facilities. Content we expect to cover includes wider data legislation related to FAIR (e.g. new revised OECD recs), relationship and linking (e.g. PID graph) and impact and FAIR metrics (e.g. how can we evaluate things together, not just separately e.g. data and

publications, not publications or data). We feel these sorts of discussions are important for raising awareness of and promoting a FAIR ecosystem.

- Gerard Coen (EOSC Synergy) the project has a number of service providers and repositories integrated in the project consortium to serve as pilots and use cases but also to expand the EOSC to new domains and broaden its capabilities. The list of repositories involvecond in the work is available here: https://github.com/EOSC-synergy Efforts to raise awareness related to FAIR are taking place within the project e.g. sharing basic onboarding materials and inviting various services and stakeholders to sessions to understand together what steps can be taken to enable FAIR for them.
- **Jakub Urban** (ARCHIVER) on an early-adapters programme. Open for organisations to participate in this project. https://archiver-project.eu/early-adopters-programme
- **Susanna Sansone** (standing in for EOSC-Life):
    - RMDkit by ELIXIR-Converge
    - FAIR cookbook by IMI FAIRplus (ELIXIR Nodes and pharmas)

16.1 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.*

16.2 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*

Rec. 17: align and harmonise FAIR and Open data policy

Policies should be aligned and consolidated to ensure that publicly-funded research data are made FAIR and Open, except for legitimate restrictions. The maxim 'as Open as possible, as closed as necessary' should be applied proportionately with genuine best efforts to share.

- **Andreas Jaunsen** (EOSC-Nordic) interview survey will be written up and discussed with policymakers by 2022.
- **Elli Papadopoulou** (NI4OS-Europe) in the process of setting up national OS cloud initiatives that will bring together key national stakeholders in academia and research in order to form synergies (approaches might vary: top down, bottom-up or hybrid) and work on producing outputs on behalf or as proposals to the Ministries, such as national OS plans and policies. FAIR are key components to suggestions made in deliverable D3.2 and in discussions, also as topics that could be collaboratively tackled via potential Task Forces created under NOSCIs (https://zenodo.org/record/4061801#.YJsEX7UzY2x). NI4OS-Europe relies on the successful model followed for 10 years now by OpenAIRE, working on the three pillars of policies-services-training on Open Access/Science and its NOADs liaising with national stakeholders, including Ministries, to realize Open Science at the national level.
- **Abigail McBirnie** (STFC, ExPaNDS) lots of work on connection between data policies at different levels and FAIR within them, e.g, facility, institutional, funder,

national, EU. We are also considering some practical issues, i.e. where practices related to Open have been simply 'transferred' over on the assumption that they also work for FAIR. Licensing is an example: while CC-By may be entirely appropriate for Open Access publications (i.e. which are copyrightable) it is not always clear that this use of CC-By can be transferred to open and/or FAIR data (i.e. which may not be copyrightable, if it is measurement data) Historically, facilities have simply applied the license they use for open publications (i.e. CC-By) to open data - but CCO may often be more appropriate for the type of measurement data produced at facilities.

- **Jakub Urban** (ARCHIVER) adapting principle "open as possible, closed as necessary".
- **Marjan Grootveld** (FAIRsFAIR) open call on policy support led to good response; FAIRsFAIR will work on this later this year to explore common elements in policies.
- **Susanna Sansone** (as FAIRsharing) work in RDA (Funders IG, Policy IG, FAIRsharing WG) harmonising policies between funders and scholarly publishers, aiming to arrive at a common template, which FAIRsharing will implement to enable comparison between policies.
- **Christian Cuciniello** (EC): working towards harmonisation of digital objects.

17.1 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.*

17.2 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.Q1 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.Q2 Any recommendations not addressed?

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