

FAIR Culture: Pillar 2

Agreements on data availability and description, data management plans, recognition and reward, policies -

Session date: 19th of May 2021

Chair: Marjan Grootveld

Rapporteur: Simon Hodson

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	Marjan Grootveld	DANS, FAIRsFAIR
2	Simon Hodson	CODATA, FAIRsFAIR
3	Abigail McBirnie	ExPaNDS/ UKRI-STFC
4	Sarah Jones	GÉANT, EOSC Future
5	Iryna Kuchma	EIFL, OpenAIRE
6	Andreas O Jaunsen	NordForsk, EOSC-Nordic
7	Angus Whyte	Digital Curation Centre, FAIRsFAIR
8	Natalie Harrower	DRI, RDA4EOSC
9	Carlos Casorrán	European Commission
10	Karsten Peters-von Gehlen	German Climate Computing Center (DKRZ)
11	Olivier Rouchon	CINES, FAIRsFAIR / EOSC-pillar
12	Antti Pursula	EUDAT CDI, DICE
13	Elli Papadopoulou	ATHENA RC / OpenAIRE and NI4OS-Europe
14	Timea Biro	DRI / NORF
15	Richard Dennis	KB /Copenhagen University Library
16	Erzsébet Tóth-Czifra	DARIAH-EU, SSHOC
17	Mustapha Mokrane	DANS/ FAIRsFAIR
18	Andreas Athenodorou	The Cyprus Institute, NI4OS-Europe
19	Ari Asmi	ENVRI-FAIR

20	Patricia Clarke	HRB-IE, FAIRSFAIR HLAC
21	Serenella Muradore Gallas	Trust-IT Services, FAIRsFAIR
22	Marialetizia Mari	Trust-IT, FAIRsFAIR

This session is about recommendations 4, 5, 6 (priority) and 18, 19, 20, 21.

Rec. 4: Develop interoperability frameworks

4.1 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 Nordic, Andreas Jaunsen: No new work. EOSC Nordic is currently implementing FAIR, following the framework provided by the FAIR assessment tools (FUJI tool and the Wilkinson et al. evaluator).

NI4OS Europe, Elli Papadopoulou, Andreas Athenodorou: NI4OS Europe opened call for experts on data standards, creating a team that can guide service providers, provide necessary services around interoperability, support data exchange in different disciplines. Supporting the long tail of science: first part of the project focussed on identifying services from different disciplines, identifying gaps. Set up teams of experts on these topics, creating best practices on scientific topics. Have produced best practices for onboarding. More in the next six months.

ExPaNDS, Abigail McBirnie: In WP2 (FAIR), we have delivered a draft metadata framework for FAIR data management across the whole experimental lifecycle. Final version to be released next year.

In relation to ExPaNDS work around federating data catalogues into EOSC, WP3 deliverable 3.1 reviewed data catalogues in photon-neutron (PaN) facilities, including undertaking a gap analysis. The WP3 ontology deliverable 3.2 is currently out for internal review. It presents several small ontologies that seek to align semantics in PaN science. The ontologies were designed with FAIR vocabularies in mind. ExPaNDS has also addressed the longer term management of the ontologies in the community.

ENVRI-FAIR, Ari Asmi: For the interoperability frameworks a lot of work is going on: the Envri-Hub idea which includes a lot of issues relating to vocabularies, use of the DCAT standard, and work regarding interoperability layers in subdomains e.g. marine science.

PaNOSC, Abigail McBirnie: PaNOSC works on data catalogues, but isn't currently doing ontology work.

SSHOC, Erzsébet Tóth-Czifra: There are interoperability problems present in SSH domain. An interoperability hub is being developed by WP3, with a first milestone report outlining main the interoperability issues in SSH. SSHOC has [mapped the DDI Codebook and CMDI to its reference ontology SSHOCro](#). SSHOC is currently discussing interoperability standards in the context of the SSHOC marketplace; more about this in June.

OpenAIRE, Iryna Kuchma: The [OpenAIRE guidelines](#) cover interoperability for data repositories/archives.

EUDAT, Antti Pursula: The EUDAT CDI provides several interoperable services. They have updated the core metadata schema for B2Share and B2Find so is now compatible with DataCite metadata schema definition v4.3.

4.2 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*

EOSC-Pillar, Olivier Rouchon: EOSC-Pillar is working towards releasing a MVP (Minimum Viable Product) for the [Federated FAIR Data Space](#) (F2DS) before the summer break. This will enable interoperability between the repositories it will harvest.

Rec. 5: Ensure data management via DMPs

5.1 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 Nordic: This topic is not really part of EOSC Nordic activities.

EOSC-pillar: DMPs are not a priority within the project. The focus is rather on developing a prototype and proof of concept for interoperability of repositories through scientific use cases.

NI4OS Europe is promoting FAIR DMPs via the ARGOS tool and through training. They have training events together with EOSC Hub and will enhance training in the next couple of months.

ExPaNDS: On DMPs we work closely with PaNOSC on creating a template of questions, based on those in the RDMOinfopool. We are looking at which questions are relevant, how do we need to change them, and are there questions we need to add? We are also looking at where the information would come from (i.e. from which systems), who would provide the information (e.g. instrument scientist, user office, researchers themselves). We would like the DMP to be populated automatically and actively as researchers move through the experimental lifecycle. We are also looking at why we are asking each question: As we don't want to ask unnecessary questions, we want to have a clear idea of how the answers to the questions will be used. Our later deliverable will be around active DMPs: How does this work in the facility, how is it completed as one moves through the experimental lifecycle.

Enrvi-FAIR: Harmonising policies relating to DMPs in policy working group.

SSHOC: We are supporting the implementation of good practice in the community through training and advocacy activities in SSHOC around DMPs. Materials from data management

planning workshop in Feb 2021 are available. There are DMP issues in relation to certain data types.

OpenAIRE: DMP's are important for training activities. We did an exercise on more than 800 H2020 DMPs. It is interesting to see what projects shared via DMPs. We had to deal with issues of licencing of the DMPs which made them difficult to share. With regard to the ARGOS DMP tool OpenAIRE applied RDA DMP standard to ARGOS so DMPs are machine actionable. Working on disciplinary aspects with a couple of research communities to implement Domain Data Protocols - DDPs. Worked with ARIADNE Plus to complement each other's work and developed a machine actionable DMP instance for archaeological data. Working with other projects representing research communities such as NEANIAS also (hydrology, and other domains). Have been approached by publishers to bridge data availability statements and dmps. The aim has been to help close the DMP publication lifecycle through integration with Zenodo (and other repositories in the future). Also working with funders such as CHIST-ERA and FCT.

DRKZ/ FAIR Champion Karsten Peters: DRKZ is trying to expand the portfolio to support DMPs.

Rec. 6: recognise & reward FAIR Data & stewardship

6.1 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 Nordic: In a task which started in Jan 2021 we study existing incentives and will encourage greater coherence on incentives. FAIR assessment is an incentive to repositories: they are now putting greater efforts into fairifying data. With respect to supporting data stewardship: NeIC has conducted data stewardship courses for the last two years; we develop competence centres for institutions.

NI4OS: what will be done in this respect will relate to national open science initiatives that have a more strategic role and their tasks include drafting/ influencing (that depends on how they are structured and operate as bottom up or top down initiatives - or hybrid even) Open Science and RDM policies and national plans that incorporate incentives and rewards based on what the global OS community follows. [Deliverable 2.2 NI4OS-Europe National OSC initiatives models](#) has some examples in the blueprint it proposes.

ExPaNDS: this is not an area on which ExPaNDS has been focusing as it is not included in our formal description of work. This said, it is becoming clear that PaN will need to consider this more in the future. Data stewardship often falls currently on the instrument scientists, but we may need to look at specific data stewardship roles in the facilities. The challenge is very much around combining the information science knowledge with the knowledge of the facility and the instruments. Certainly, if nothing else, there remains the open question of how to recognise and reward the effort that instrument scientist facility staff do put towards data stewardship. At the moment, this is not really addressed, and it probably fair to say that instrument scientists are just expected to take this on as yet more additional work. There isn't really the widely promoted concept of formal recognition for that.

ENVRIFAIR: In a sense this is part of the culture in infrastructures; it is a culture question. There is also the issue of authorship of the data: how to balance the observation made by

the facility and the way the data are put to use by the scientist. This affects the issue of recognition.

SSHOC: There is a high level of awareness of the need for better recognition, but a feeling also that issues of reward and recognition fall out of the scope of the project. There is an issue of how community curation will be managed and rewarded in the SSHOC market place, badging is under consideration. Risks of gaming of badging system.

DRKZ: There is an issue of whether the data is given a DOI and whether it counts as an output, and therefore included in / recognised as a scientific output. In the climate computing centre, a DOI is only given for quality assured data complying with the [submission guidelines](#). By the way, the very first DataCite DOI for data was given by the WDCC. The process until the submitted data is finally archived may take several weeks of intense back-and-forth communication between WDCC staff and the data providers. We are working on approaches to make the quality of archived data visible for data reusers, e.g. within the [AtMoDat project](#).

Rec. 18: cost data management

18.1 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.*

ExPaNDS: One of the themes of the data management framework is responsibility for costs (but at a quite high level). In the DMP template, we have quite a few questions in relation to costs, but this presents a challenge to facilities, as it is difficult to calculate cost information.

OpenAIRE: Has guidance and infographics developed in partnership with some institutions that are OpenAIRE AMKE members. These are reusable resources that are being used in trainings and relevant activities.

DRKZ: We don't have a cost model, but there is increasing demand from projects that have funding for DMP FTEs. Increasingly these are placed with the WDCC to support the project data management.

Rec. 19: select and prioritise FAIR digital objects

19.1 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 Nordic: Recommend and promote the concept, support the idea of fairification. Again this relates to the use of assessment tools.

EOSC Pillar: Work on the impact that the FDO could have on services is in the evaluation phase. FDO will be considered more in phase two of our MVP. We are evaluating the concept and how it will impact the services, how it could be implemented.

ExPaNDS: There is the question to which data we assign a PID and in what stage of the process: To the 'raw' data or later in the process? We have so much data, which goes through so many processes through the data lifecycle: this raises questions about what to keep and what needs to have PIDs etc. The PAN community is starting to think more about which data to select for FAIR, also involving PaNOSC. Workshops will consider what data needs to be curated to what level, and what data needs to be retained in the long term, i.e. how can we be selective in the data we make FAIR.

Rec. 20: deposit in Trusted Digital Repositories

20.1 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 Pillar: 'EDTR' is the long term preservation platform deployed at CINES. It is part of the EOSC catalogue of services.

NI4OS: One of the objectives of the project is to onboard repositories. In discussions with repository managers, we inform them about the FAIR principles, about CTS. So far none have applied for CTS.

ExPaNDS: PaN facilities' catalogues are federated through EOSC. There are two related strands of work: 1. Creation of a PaN search API (completed): this allows to search across all facility catalogues in EOSC. 2. Metadata harvesting using OAI-PMH into B2Find and OpenAIRE.

ENVRI-FAIR: We recommend to have some level of certification. A number of the facilities have secondary data storage.

DICE, Antti Pursula: There is a task on long-term archiving of data, in which we are drafting a strategy / vision on long-term preservation of EUDAT services.

DRKZ: WDCC has been certified with CTS for a few years now. The global community can ask whether they can archive their data with WDCC. This is generally free of charge, unless there is a particularly high volume or demand.

FAIRsFAIR, Mustapha Mokrane: A lot of work has been done for developing the maturity model for FAIR certification for repositories. We are also aligning the concepts contained in FAIR and TRUST principles.

Rec. 21: incentivise reuse of FAIR outputs

21.1 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.

NI4OS: The NI4OS-Europe training catalogue collects resources that are available to everyone. Resources can be reused and they can also feed into the EOSC training catalogue once it is ready.

ExPaNDS: We are federating into EOSC to encourage the findability of data generated at PaN Research Infrastructures. Hopefully this will lead to greater reuse.

ENVRI-FAIR: This incentivisation can also happen through services which make reuse easier. Sharing of notebooks, workflow methods between users is important in this respect.

EC, Carlos Casorrán: There is a relation to DMPs, as DMP authors are expected to provide information of what data the data creators think will have reuse value.

OpenAIRE: In ARGOS, we try to normalise the different types of data and the criteria for reuse. E.g. for sensitive data we try to understand what researchers and funders need to know. ARGOS draws information from the OpenAIRE API and creates links with different research entities and outputs included in the OpenAIRE Research Graph. We introduced a reused data section in some templates (e.g. Science Europe, CHIST-ERA and H2020) with fields that gather information relevant for data that are being reused in the research, e.g. link to corresponding DMPs (if any, to help with providing additional information), data repositories to know where reused data resides/ are preserved, PIDs/name of datasets to know which particular datasets from those repositories are being reused etc.