Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D3.7 Report on integration of metadata catalogues

Work Package	WP3
Lead Author (Org)	Simon Lambert (STFC)
Contributing Author(s) (Org)	Ricarda Braukmann (DANS-KNAW), Joy Davidson (DCC), Eva Méndez (UC3M) Marina Sánchez (UC3M)
Due Date	30.11.2021
Date	30.11.2021
Version	1.0
DOI	https://doi.org/10.5281/zenodo.5744912

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.1	22.07.2021	Simon Lambert	Initial structure proposed
0.2	09.09.2021	Simon Lambert	Revised structure
0.3	27.10.2021	Joy Davidson, Ricarda Braukmann, Marina Sanchez, Eva Mendez, Simon Lambert	Final ToC, structure, initial content
1.0	4.11.2021	Joy Davidson, Ricarda Braukmann, Marina Sanchez, Eva Mendez, Simon Lambert	Complete content ready for internal review
1.1	29.11.2021	Simon Lambert	Finalised after internal review by Sarala Wimalaratne and Hanna Koivula and checking by interviewees
1.0	30.11.2021	Simon Lambert	Submitted to PCO for publication to Zenodo

Disclaimer

FAIRsFAIR has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no. 831558 The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

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

Abbreviations and Acronyms

API	Application Programming Interface
CDC	CESSDA Data Catalogue
CESSDA	Consortium of European Social Science Data Archives
CMDI	Component Metadata Infrastructure
CMM	CESSDA Metadata Model
CODATA	Committee on Data of the International Science Council
CRIS	Current Research Information System
CSW	Catalogue Service for the Web
DCAT	Data Catalog Vocabulary
DDI	Data Documentation Initiative
DDI-CDI	DDI Cross Domain Integration
DOI	Digital Object Identifier
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
FDO	Feature Data Objects
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
NGSI-LD	Next Generation Service Interfaces – Linked Data
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OWL	Web Ontology Language
RDA	Research Data Alliance
SDTL	Structured Data Transformation Language

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

SDMX	Statistical Data and Metadata Exchange
SOSA-SSN	Sensor, Observation, Sample, and Actuator (within Semantic Sensor Network ontology)
SSH	Social Sciences and Humanities
W3C	World Wide Web Consortium
XML	Extensible Markup Language

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

Executive Summary

The preceding work of FAIRsFAIR in metadata catalogue integration was presented in deliverable D3.6 "Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake". The challenge was recognised to be facilitating cross-disciplinary data discovery, with metadata catalogues having a key role to play in that. The deliverable described the complex challenge of facilitating data discovery, and outlined the many approaches and metadata standards in use. It analysed the distinctions between data catalogues and metadata catalogues, and between domain-independent/agnostic metadata and domain-specific metadata. It analysed several domain-agnostic metadata standards and proposed a pilot activity to assess two particular domain-agnostic metadata standards, namely DCAT 2 and DDI-CDI.

Subsequent work is reported in the present deliverable. A series of workshops were held with stakeholders representing research communities and other interested parties. The rather ambitious pilot programme envisaged in D3.6 was reduced in scope to something more realistic and achievable: the pilot focused on exploring some of the practical implementation challenges that currently hinder the reuse of existing metadata mappings, to get a sense from different perspectives about the suitability of DCAT 2 and the planned improvements for DCAT 3 as an integration solution, and to consider some of the additional factors that should be addressed. It was further recognised that DCAT and DDI-CDI are not competing standards but complementary, with DCAT concerned with discoverability and DDI-CDI with interoperability.

The FAIRsFAIR team worked with representatives from two of the thematic cluster projects (SSHOC and PaNOSC) and two service providers (B2FIND and OpenAIRE) to assess the feasibility of DCAT 2 from both the domain specific and aggregator perspectives. The intention is not to be prescriptive, but to energise the community and set the perspectives. In all cases a positive attitude to DCAT was revealed, with no obstacles in principle to its adoption. The key factor is the existence of demand in a world of constrained resources. Other issues also arose and are presented in this report, such as the implications of the differences between research domains, requirements for sustainability (in a broad sense), and an emergent desire for a central documented collection of metadata mappings.

Work will continue in the remainder of the FAIRsFAIR project to progress community dialogue in these areas and to set up the living resources such as the collection of mappings.

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

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

1. Introduction and context

1.1 Background

It is widely accepted that sharing and reuse of research data is fundamental for facilitating and accelerating the research process and especially interdisciplinary research. The first step to reuse is of course discovery of suitable data—the F (Findable) of FAIR. Data that is made available for sharing by its originators needs to be exposed to discovery, and in practice that means including it in appropriate repositories that expose its metadata for human search or automated harvesting by machines. The metadata itself will reside in a catalogue: the distinction between data catalogue and metadata catalogue is somewhat academic, as the preceding deliverable from FAIRsFAIR D3.6 (Méndez et al., 2020) recognises: "A metadata catalogue contains information about a dataset, with a persistent identifier, that enables the placement of that dataset into context. A data catalogue can be the same as a metadata catalogue, with the exception that it also contains the data that the metadata describes." Metadata catalogues exist at several different levels. Major search engines are for many users the main entry point and function in that way as metadata catalogues. Metadata integrators such as B2FIND and OpenAIRE are discovery portals covering multiple domains, while domain-specific repositories and their catalogues provide a deeper set of metadata relevant to the domain.

The capacity for discovery of datasets therefore depends on the suitability and quality of the metadata in the catalogue, though there are widespread deficiencies in this respect. The report on the EOSC Interoperability Framework (EOSC Executive Board, 2021) presents an incisive analysis of the shortcomings of the present situation, with multiple metadata standards depending on the domain, legacy metadata not updated, poor conversion to the generic standards used by aggregators, lack of data types and controlled vocabularies, and so on. The EOSC Interoperability Framework makes its own recommendation of a minimum metadata set to describe metadata records for FAIR data.

Effective data discovery is not only a matter of adequate standards: it is performed by researchers with particular needs in the context of their work. A recent study (Gregory, Cousijn & Groth, 2019) conducted interviews with researchers on data needs, search practices and evaluation behaviours. It recognises that searching for data involves interactions between technical and social aspects, and indeed refers to the "inherently social nature of data search", yet one of the conclusions is that "Descriptions about data can be as important as the data themselves".

The situation has, not surprisingly, attracted the attention of the Research Data Alliance (RDA) through the Research Metadata Schemas Working Group[1]. The starting point was the observed growth of *schema.org*, but with the need for a more systematic and controlled approach in the case of research data. The Working Group has the objective to identify and bridge gaps in existing schemas commonly used for research data and provide guidelines for communities, and in fact provides crosswalks to *schema.org* as well (linked from the Working Group's homepage). The EOSC Interoperability Framework also offers a comprehensive crosswalk (linked from the report on the

[1] https://www.rd-alliance.org/groups/research-metadata-schemas-wg

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

Framework)—though the profusion of crosswalks with differing motivations and starting points can itself be a source of perplexity.

At the European level, EOSC is of course highly motivated to enable discoverability of its assets. There is a technical specification[2] of metadata cataloguing and indexing as a common service of EOSC. Metadata cataloguing and indexing is taken to include uploading and indexing of metadata records in the metadata catalogue, to enable data discovery and access. A simple overview of the process is presented as a "high-level service architecture". Metadata mapping is seen as a responsibility of the overall service provider ("normalisation, homogenisation and mapping of the specific community standards onto a generic, common and unified metadata schema should be performed").

The EOSC Portal is one of the key services of the so-called "federating core". Integration of entire existing catalogues has emerged as an important requirement, and the projects EOSC Enhance[3] and NI4OS-Europe[4] are working together on a pilot with this aim, though the emphasis is on the technicalities of the onboarding process, profiles and APIs rather than discoverability *per se*.

The research infrastructures represented in the five cluster projects launched in 2019 (Gotz et al., 2020) are fundamental through their links with research communities and their management of large amounts of data. The EOSC Enhance project has conducted a study of their existing research data cataloguing efforts (Goble & Juty, 2020) with a view to encouraging adoption of common metadata interoperability guidelines. If there is one overarching conclusion, it is the recognition of the heterogeneity of terminology, concepts and practices and outlooks in the sphere of data management. The infrastructures recommend standards for discoverability and low-level interoperability, but the report makes the trenchant observation that "Interoperability for its own sake is unproductive and wasteful; the clusters expend effort on interoperability between metadata and their catalogues only where and when it makes sense (through real examples) and brings benefit."

1.2 The perspective of FAIRsFAIR

FAIRsFAIR is a practical project, aiming to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. It is closely aligned with EOSC, and the expectation is that outputs of FAIRsFAIR will contribute to embedding FAIR practices into EOSC. One of the project's tasks is on metadata catalogue integration, defining what is needed to enable data repositories and service providers to support FAIR data in terms of general practice and the specifics of metadata handling and integration to facilitate cross-disciplinary data discovery.

The motivation is to achieve increased production and use of FAIR data, and the high-level use case is simply "to make our metadata catalogue available"; the beneficiary being the repository service, rather than the end-user researchers themselves, since their own use cases and modes of working can be many and various—"It is not enough to think of data users as researchers in a discipline with

[2] https://wiki.eosc-hub.eu/display/EOSCDOC/Metadata+Management+and+Data+Discovery
[3] More information on the EOSC Enhance project: https://eosc-portal.eu/enhance
[4] More information on the NI4OS-Europe project: https://ni4os.eu/

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

fixed practices" (Gregory, Cousijn & Groth, 2019). Nonetheless the preceding deliverable D3.6 did provide some outline motivating use cases relating to different domains but with cross-domain aspects, for example, connecting clinical datasets indicating infection rates to datasets which contain geospatial and population information.

D3.6 examined the practices of the five ESFRI clusters (ENVRI-FAIR[5], PaNOSC[6], ESCAPE[7], SSHOC[8], EOSC-Life[9], as in the EOSC Enhance deliverable already cited) and proposed a pilot activity to test the (meta)data catalogue integration. It rejected the idea of developing and applying common standards across domains and institutions, and drew a distinction between *domain-agnostic metadata* and *domain-specific metadata* schemas, though it should be remembered that some metadata elements, such as spatial or temporal delimitations, are independent of any particular domain but do not apply to all datasets. DCAT 2[10] was selected as the domain-agnostic standard to be evaluated from the perspectives of both repository services and aggregators of metadata. DCAT 2 was selected in part to reflect the fact that a number of current EOSC related initiatives are also considering the potential value of DCAT including ENVRI-FAIR and EOSC Pillar (Grootveld et al., 2021). In addition, the aggregator service re3data recently outlined plans to update its API to enable repositories exposing their catalogues in DCAT to be supported (Wimalaratne & Ulrich, 2021). As stated in D3.6: "The findings of this pilot will be presented in D3.7 […] and, based on the outcomes, FAIRsFAIR will provide guidance to assist other repositories to improve the FAIRness of their data collections through integration of their metadata catalogues."

The remainder of this report is structured as follows:

- description of the methodology and overview of the pilot activities
- analysis of the interviews that were conducted as part of the pilot
- survey of some other relevant work on data discovery and interoperability from different perspectives
- conclusions including further investigation and engagement to be done in the remainder of FAIRsFAIR and beyond

[5] https://envri.eu/home-envri-fair/
[6] https://www.panosc.eu/
[7] https://projectescape.eu/
[8] https://www.sshopencloud.eu/
[9] https://www.eosc-life.eu/
[10] https://www.w3.org/TR/vocab-dcat-2/

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

2. Methodology and pilot activities

As can be seen in Figure 1, the pilot as outlined in the preceding deliverable D3.6 envisaged an ambitious programme of working with domain specific participants to carry out a mapping between a relevant domain specific standard and DCAT 2. In parallel, B2FIND had considered working to implement a DCAT reader and the necessary mapfiles to support the harvesting of metadata catalogues in this format.

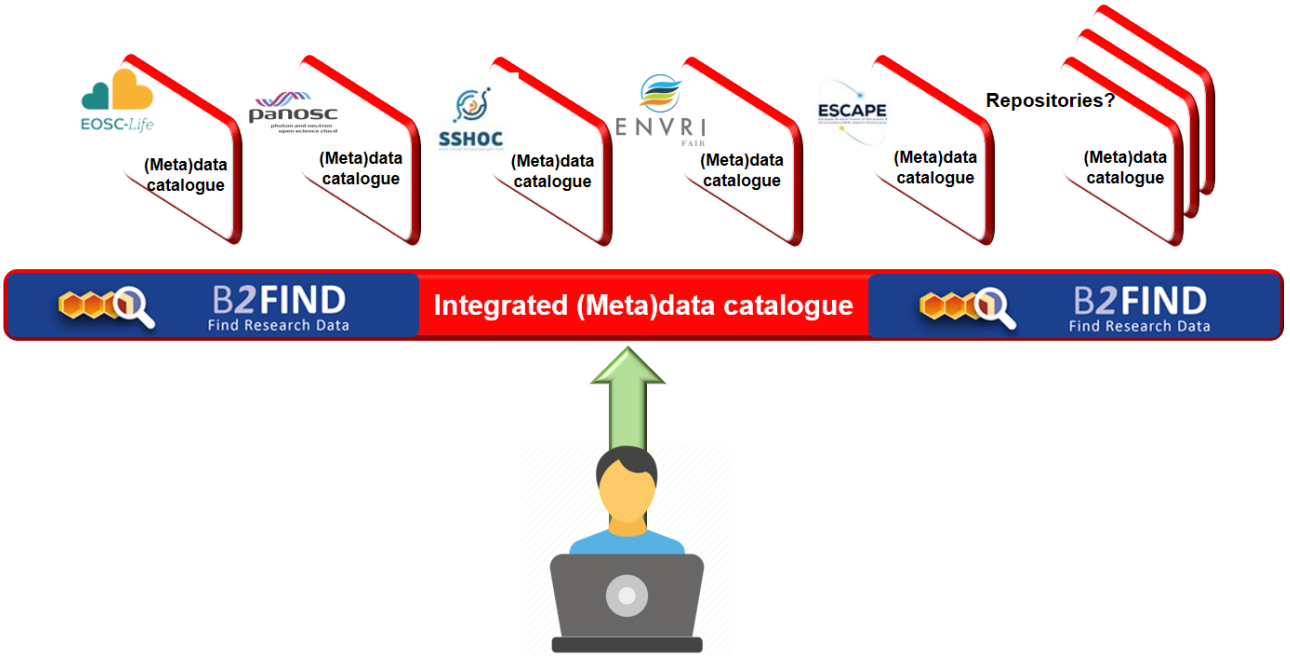

Figure 1: FAIRsFAIR conceptual proposal of (meta)data integration catalogue

However, it became clear as work began during the early stages of the pilot that there are many metadata mappings that already exist providing crosswalks to DCAT and other domain-agnostic standards such as *schema.org*. In addition, as work progressed the required effort and resources needed to develop a DCAT 2 reader within the B2FIND service were ultimately not justifiable as there is currently a lack of demand for such functionality from the repositories that provide content. To avoid duplicating effort on metadata mapping and to reflect the lack of demand for DCAT 2 harvesting, the pilot shifted focus. Instead, the pilot focused on exploring some of the practical implementation challenges that currently hinder the reuse of existing metadata mappings, to get a sense from the discipline specific and aggregator partners about the suitability of DCAT 2 and the planned improvements for DCAT 3[11] as an integration solution, and to consider some of the additional factors that should be addressed.

An important event leading to this shift in focus was a workshop organised by FAIRsFAIR in March 2021, the third in the series dedicated to metadata catalogue integration. The participant list and programme of the workshop are reproduced in Annex 2. An in-depth presentation and discussion was held on DDI-CDI (see section 4.2 below), which clarified the relationship of this

[11] https://www.w3.org/TR/vocab-dcat-3/

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

interoperability-oriented standard with discovery-oriented standards such as DCAT, and led to the recognition that these two standards should not be treated as "competing" alternatives but as complementary. A further outcome of the workshop was a widespread desire for systematic documentation of the existing crosswalks[12]. However, though there was clear interest in principle in the aims of the pilot, ultimately and for reasons just explained the initial ambitions had to be circumscribed.

Therefore, over the summer of 2021, FAIRsFAIR carried out a small-scale pilot to assess the potential value and current limitations of DCAT 2 to support discoverability of metadata catalogues by aggregators and to help increase the findability and reusability of data across discipline specific repositories. During the pilot, the FAIRsFAIR team worked with representatives from two of the thematic cluster projects[13] (SSHOC and PaNOSC) and two service providers (B2FIND and OpenAIRE) to assess the feasibility of DCAT 2 from both the domain specific and aggregator perspectives. The intention is not to be prescriptive, but to energise the community and set the perspectives.

The approach was pragmatic: two very different domains were chosen, concentrating on mapping to DCAT 2 within the context of the wider priorities and concerns around metadata catalogue integration. From the Social Sciences and Humanities (SSH) community, a representative from CESSDA, the Consortium of Social Science Data Archives was interviewed. From the photon and neutron community, a representative from the PaNOSC (Photon and Neutron Open Science Cloud) project was interviewed. This was complemented by two interviews with aggregators (B2FIND and OpenAIRE) for the other side of the story. The interviews were based on a concise but comprehensive questionnaire (see Annex 1).

The revised methodology is depicted in Figure 2.

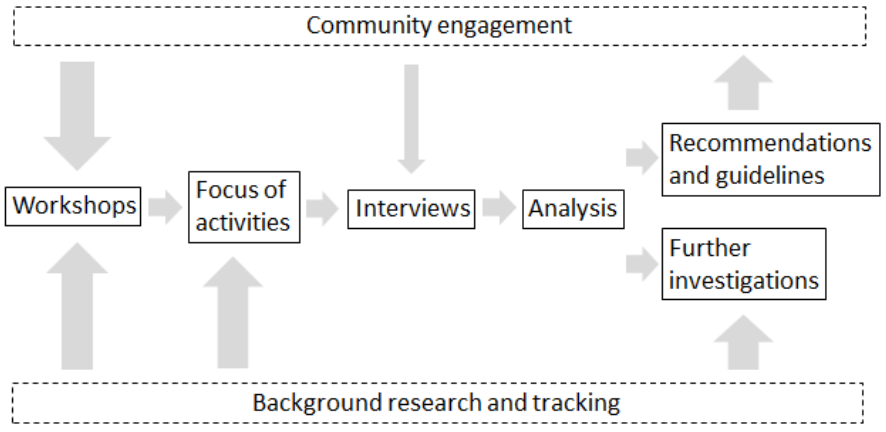

Figure 2: Revised methodology for the pilot activity

[12] In this context, the words "crosswalk" and "mapping" between metadata schemas tend to be used synonymously (Zeng and Qin, 2016).
[13] https://www.eosc-hub.eu/eosc-hub-and-esfri-cluster-projects

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

3. Pilot study of metadata catalogue integration with DCAT

In this section, we provide a summary of interviews carried out with selected stakeholders to gain a sense of the potential suitability of DCAT to support integration, current limitations, and the initial feedback on the revisions of the vocabulary outlined in the W3C working draft of DCAT version 3 released in May 2021. The interviews also aimed to identify other practical issues that hinder metadata catalogue integration.

3.1. Introduction to the interviews with stakeholders

As noted earlier, this small pilot involved carrying out interviews with representatives from two of the thematic cluster projects (SSHOC and PaNOSC) and two service providers (B2FIND and OpenAIRE) to assess the feasibility of DCAT 2 from both the domain specific and aggregator perspectives. In this section, we provide further information about the interview participants and the initiatives they represent.

Table 1. Summary of interviews

Initiative	Stakeholder Group	Interviewee
CESSDA	Social Science Data Archives, leader of the SSHOC cluster project	Mari Kleemola, FSD Finnish Social Science Data Archive
PaNOSC	Photon/Neutron Community	Tobias Richter, European Spallation Source
B2FIND	Aggregator service provider	Claudia Martens and Anna-Lena Flügel, Deutsches Klimarechenzentrum
OpenAIRE	Aggregator service provider	Andreas Czerniak and Jochen Schirrwagen, Bielefeld University

CESSDA

The Consortium of European Social Science Data Archives (CESSDA) is one of the ERICs active in the field of the Social Sciences and Humanities (SSH). CESSDA is involved in various EOSC projects and activities and is the lead partner in the Social Sciences and Humanities Open Cloud (SSHOC)[14] project, the ESFRI cluster project representing the SSH community. CESSDA operates the CESSDA Data Catalogue (CDC)[15], which collects metadata from various European data archives who store and curate social science datasets. The CDC currently offers metadata from more than 33000 studies in various languages. The metadata is organized according to the CESSDA Metadata Model

[14] SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782.
[15] The CDC is available on the CESSDA website: https://datacatalogue.cessda.eu/

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

(CMM)[16] and the CDC profile[17], an implementation of the Data Documentation Initiative (DDI)[18] standard which is widely used in the social sciences, although other specialised standards are used in the SSH subdomains, e.g. linguistics (CLARIN) has CMDI[19].

The CDC is not currently machine-harvestable, but is rather an interface for human use in querying. However there are plans for machine access through the CESSDA Aggregator Project, which will provide metadata records via OAI-PMH, for example in Dublin Core and DDI Codebook 2.5. Individual CESSDA providers (i.e. data archives) usually make their metadata machine-accessible.

CESSDA makes extensive use of multilingual controlled vocabularies[20] and provides a multilingual thesaurus[21] to assist with data discovery and a special controlled vocabulary for topic classification.

PaNOSC

There is an important body of researchers using sources of photons and neutrons to examine the structure of materials. They conduct their investigations at large-scale facilities that provide these sources, typified by the European Spallation Source, Institut Laue-Langevin in France and ISIS in the UK. Consequently this type of research is also known as facilities science, and it has developed its own particular practices and workflows. There is a well developed data/metadata format called NeXus, which in addition to the detector data includes sample, instrument and scientific metadata.

The community is very active in European projects: PaNOSC[22] is one of the ESFRI cluster projects, developing and providing services for scientific data and connecting these to the EOSC. Another current project, ExPaNDS, concentrates on data sources and data analysis services for facilities science. The SciCat project is under development by several photon/neutron sources as a comprehensive scientific data management environment with an orientation to FAIR data.

Data sources are already harvested by B2FIND, and there is a mapping from NeXus to Dublin Core.

B2FIND

B2FIND[23] is a discovery portal for research outputs such as data collections that was developed within the EUDAT projects and is now a basic service of the EUDAT Collaborative Data Infrastructure [24] and a service of EOSC. It is multidisciplinary and harvests and aggregates metadata from community repositories as well as EUDAT data centres, and allows free text and faceted search. It is itself harvested by OpenAIRE.

[16] The CMM is available on Zenodo: https://doi.org/10.5281/zenodo.4751455
[17] CDC metadata profile: https://cmv.cessda.eu/profiles/cdc/ddi-2.5/1.0.4/profile.xml
[18] Website of the DDI: https://ddialliance.org/
[19] https://www.clarin.eu/content/component-metadata
[20] https://vocabularies.cessda.eu/
[21] https://elsst.cessda.eu/
[22] www.panosc.eu
[23] http://b2find.eudat.eu
[24] https://eudat.eu/eudat-cdi

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

B2FIND has its own metadata catalogue using the EUDAT Core metadata scheme[25], used to harmonise and aggregate the metadata it harvests. Harvesting is mostly of XML or JSON records via OAI-PMH or CSW but also via APIs; B2FIND works with individual repositories to access metadata in the most straightforward and practical way. Currently the following metadata standards are supported (i.e. already implemented as "readers", so that records using them are easy to integrate): Dublin Core, DataCite (generic); ISO 19139/19115, FGDC (geographic metadata, thematic); and some domain-specific standards such as CMDI for linguistics. Currently not supported are DCAT and DDI/DDI-CDI.

OpenAIRE

OpenAIRE[26], developed through a series of projects funded by the European Commission, has developed into a broad collection of services to support Open Science. The services are aimed at stakeholders including researchers, content providers and research funders. Currently a total of 27 services are listed.

For the purpose of this study, the OpenAIRE Explore Portal is the most relevant service. This is a discovery service based on OpenAIRE's open scholarly communication graph that includes all research and scholarly activities across the research lifecycle. The graph is created bi-monthly by aggregating, cleaning, transforming and inferring content retrieved from a wide range of validated Open Access providers[27]. These include institutional and thematic repositories, data repositories, journal platforms (JATS), metadata catalogues and CRIS (CERIF).

Harvesting is done mostly via OAI-PMH interfaces with repositories directly and mostly using Dublin Core or DataCite metadata. OpenAIRE's own OAI metadata prefix *oai_openaire* is used for metadata records following the OpenAIRE Guidelines Application Profile v4[28]. It is also possible to gather data from REST APIs or FTP.

3.2. Analysis of the interviews

3.2.1 The prospects for DCAT

The interviews covered the suitability of DCAT 2, any perceived challenges concerning its use, and views on the extensions being introduced in DCAT 3 to feed back to the DCAT 3 working group.

Although none of the systems examined is currently exposing or harvesting DCAT metadata, there was universal agreement that the model is a very suitable one and would map well to current practice with no major obstacles foreseen. The high-level concepts such as Catalog, CatalogRecord, Resource and Dataset were regarded as natural and in some cases useful extensions to the models currently used. DCAT was regarded as a very rich generic approach.

[25] Introduction at https://eudat-b2find.github.io/schema-doc/introduction.html
[26] https://www.openaire.eu/
[27] The OpenAIRE content acquisition policies are reported in (Becker et al., 2018)
[28] https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/latest/

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

There were some questions from the aggregator side about extensions that are not currently included in DCAT, such as handling scientific instruments—this is not a purely domain-specific consideration, since many domains make use of instruments requiring identification and referencing. OpenAIRE supports person/author, organisation, funder/project elements which are not covered by DCAT. (See also section 3.2.4 on difficulties with handling some domain-agnostic elements in general.)

DCAT 3 is currently under development[29] through the regular procedures of W3C. A number of requirements have been identified and extensions proposed. Of particular interest are the introduction of terms and guidelines on versioning and dataset series. Versioning will allow DCAT to describe versions resulting from a revision, i.e., from changes occurring to a resource as part of its life-cycle. Dataset series refers to interrelated datasets that are published separately, such as datasets for successive years or different countries.

The enhancements concerning versioning and dataset series envisaged in DCAT 3 were seen as valuable, as they are not uniformly well handled, so there is interest in their standardisation in this way. Dataset series are certainly important for the social sciences. There is a connection to provenance, as also for versioning, where it might be important to specify how a new version supersedes an old.

3.2.2 Metadata mappings

Metadata mappings are a crucial component for supporting findability and interoperability across heterogeneous data collections. As noted in a recent paper on making the EOSC work (Jones et al., 2021),

> "We need an EOSC data search engine, comprehensive and simple enough as Google. In order to achieve this we need to federate all the individual data catalogues under a common search API and aggregate the information from search requests into something meaningful. Naturally that means in-depth community work on mapping disciplinary metadata standards and ensuring semantic interoperability so we can integrate content across repositories. There is also a need for coordination and agreement on broader aspects of data structure and provenance to enable reuse, not just discovery."

During our interviews and desk research, we found that mappings from other metadata standards to DCAT 2 are already available in some cases (i.e. from DDI to DCAT 3 based on study descriptions) and were otherwise not seen as a large obstacle. Generally, the domain specific interviewees thought that their currently used standards could map well to DCAT 2.

One issue that was mentioned in the discussions around metadata mappings was that the mappings and crosswalks that are being created are not consistently documented, centrally shared and that the governance of these resources varies. Finding existing mappings and following ongoing work is therefore seen as a major challenge. A more coherent approach to documenting and sharing

[29] https://www.w3.org/TR/vocab-dcat-3/

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

mappings through a central registry was discussed and the question was raised whether registry services such as FAIRsharing and RDA initiatives could play a central role. In addition, the aggregators B2FIND and OpenAIRE reported that thus far requests to harvest DCAT 2 have been very limited and crosswalks to (new) community standards are typically being developed on a case-by-case basis based on demand.

3.2.3 Sustainability issues

Metadata catalogue integration is not a once-and-for-all activity. There is a complex interplay between stakeholders that changes over time. New data sources become available, and old ones might change their practices; the governance of standards leads to extensions, reworkings, and in some cases backwards incompatibilities that require adaptation. These are aspects of sustainability, in the sense of the capacity to keep an operation or activity continuing into the future[30].

One of the key findings from the interviews is that there must be demand for use of a particular standard in order to justify prioritisation of the work needed to accommodate it. Theoretical value is not enough on its own. Aggregators of metadata engage in a dialogue with their communities, and generally work with (meta)data providers in straightforward and practical ways so as to ease the tasks that the providers must carry out. The aggregators must adapt themselves to what is on offer, which might encompass a highly diverse range of sources, as for OpenAIRE. The need for pragmatism tends to override idealistic considerations.

On the side of the providers, there are sustainability issues around legacy metadata and its variable quality. Social sciences archives contain data collected a long time in the past whose metadata might now be inadequate or obsolete. Resources for improving metadata quality and legacy metadata are desirable. Work can only be partially done at the level of the aggregator: more is needed to improve metadata quality and compliance at the local archives which deliver the data.

Finally, the big question of funding cannot be avoided. Too often mature services are supported through successive project-based funding, which while welcome does not necessarily match the needs of the service. On the data provider side, large institutes with copious resources can support metadata choices, but small institutes with few resources will be much more constrained in what they can offer. Anything that can assist in practical decisions and actions is welcome, such as the suggested central registry of metadata mappings.

3.2.4 Domains, communities and their diversity

The simple distinction between “domain-agnostic” and “domain-specific” metadata standards conceals a great complexity, though even if taken at face value it raises questions about the nature of the interface or handover between the two.

In the case of photon/neutron facilities science, the researchers come from their own disciplines and are using the techniques offered by the photon/neutron sources to conduct their investigations. The material under investigation might originate in a field with its own metadata conventions

[30] The FREYA project on infrastructure for persistent identifiers proposed that there are three facets of sustainability: maintainability, adaptability and desirability (Lambert, 2018).

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

(history, geology, metallurgy, etc.). The social sciences also encompass a wide range of subdomains with their own practices. On the other hand there are some aspects that are in common with other fields: the use of specialised instruments requiring persistent identifiers, for example, is certainly not restricted to photon/neutron science.

In any case, the distinction between domain-agnostic and domain-specific is blurred. There are many attributes that are certainly not specific to a particular domain, yet not applicable to all datasets—spatial and temporal coverage being obvious examples. Even though these might be considered essential across domains, the diversity of needs can make it very difficult to implement in a useful way: for example the vast orders of magnitude differences in time intervals between nuclear physics, archaeology and geology.

From the aggregators' perspective, the key to avoiding shifting and unclear requirements is community acceptance of standards, which in turn relies on good governance, for example in the creation and maintenance of community profiles. In principle, richness is the goal (as much metadata as possible), with inclusion of already existing "standards" or closed vocabularies and persistent identifiers wherever possible. Moreover, there might be benefits from linking with other domains using similar research techniques, as in the example of instruments given above.

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

4. Other perspectives on metadata catalogue integration: standards, mappings and use cases

4.1 Revisiting the challenges in the context of FAIR

In general terms, FAIR implementation has been rather unbalanced: a considerable focus on discoverability and persistent identification (Findability); some discussion about data assessment, integration, and harmonization (Interoperability and Reuse); but very little on Accessibility. Thus, while DCAT has great potential as a standard to support Findability, it must be linked to other standards and tools to also support Accessibility, Interoperability and Reuse.

All metadata schemas and element sets aspire to become a standard but not many become more than a tacit agreement within a specific domain. When it comes to cross-domain data discovery, data documentation and metadata are crucial. The gap between metadata standards at the global level and description needs at the local level is one of the challenges to make research data FAIR. The idiosyncrasies of different disciplinary fields create data complexities that are often unique to a field, while a metadata standard for research data must be general enough to accommodate metadata description needs from all fields under a broader disciplinary domain. The fact that almost all data repositories use metadata schemas modified from one or more standards is the best illustration of such a gap (Tompkins et al., 2021). Many institutions and projects have created initiatives, committees and groups to provide support for a wide variety of research data activities, including the analysis of elements and properties to create metadata mappings among different schemas, while others are analysing use cases to explore cross-domain integration of research data/metadata catalogues (as described below in the case of DDI-CDI).

Multiple metadata schemas exist within the various subject disciplines[31], although many of the main standards collect similar properties and elements to describe and document data. The correspondence between metadata elements among different metadata standards is a usual way to achieve interoperability at schema level, and is referred to as schema mappings or “metadata crosswalks” (Haynes, 2018; Zeng and Qin, 2016) .

The current landscape of metadata standards for research data is as rich as it is complex, complicating FAIR data sharing within domains. However, replacing the old approach with a unique single metadata schema is not an option. Thus, a cross-domain “lingua franca” (as easy as possible) is needed, based on existing resources in domain-agnostic contexts and unifying approaches from different domains and institutions. Among all the approaches for implementing FAIR, a common idea is to establish a set of standards for supporting the exchange of metadata. What is becoming clear through this pilot and other related initiatives is that there is a potential role for the adoption of DCAT to support findability and Data Documentation Initiative Cross-Domain Integration (DDI-CDI) specification[32] to support interoperability between disciplines.

[31] FAIRsFAIR presented a clear approach to metadata standards and specifications in section 4.3 of D3.6.
[32] https://ddialliance.org/Specification/ddi-cdi

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

4.2 Selecting and linking standards

One approach to the problem of linking together standards for Findability and Interoperability has recently been made by CODATA, which, in partnership with the DDI Alliance, carried out a European Open Science Cloud co-creation project to examine the potential uses and applications of the draft DDI-CDI specification to support interoperability within the European Open Science Cloud. DDI-CDI "provides a model for working with a wide variety of research data across many scientific and policy domains. It provides a level of detail which supports machine-actionable processing of data, both within and between systems, and is designed to be easily aligned with other standards"[33]. The resulting report was launched at a virtual workshop held in June 2021 and illustrated the potential value of using a combination of DCAT to support findability and DDI-CDI to support interoperability (as shown in Figure 3).

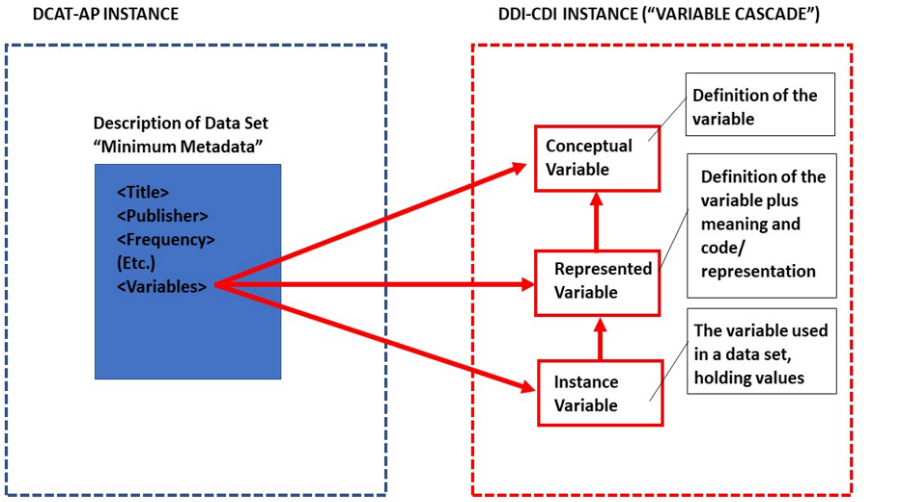

Figure 3 (Envisioned) DCAT-AP and DDI-CDI as presented at The Role of DDI-CDI in EOSC: Possible Uses and Applications report launch workshop[34]

Communities of Practice Use Cases defined during "Interoperability for Cross-Domain Research" Dagstuhl Workshop 2021[35] are another way to find cross-disciplinary metadata integration. This workshop was the result of the work of DDI in cooperation with CODATA and was based on the conclusions of two previous Dagstuhl Workshops in 2018 and 2019[36]. Its main aim was to analyse the issues and limitations towards the sharing of FAIR data across and within domains and to explore its possible solutions to provide guidance for research communities. The workshop focused

[33] DDI – Cross Domain Integration: Introduction. https://ddi-alliance.atlassian.net/wiki/download/attachments/860815393/Part_1_DDI-CDI_Intro_PR_1.pdf?version=1&modificationDate=1584721519637&cacheVersion=1&api=v2

[34] https://codata.org/initiatives/decadal-programme2/ddi-cross-domain-integration/ddi-cdi-and-eosc/

[35] https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/2677112836/2021+Interoperability+for+Cross-Domain+Research+Use+Cases+for+Metadata+Standards

[36] https://codata.org/initiatives/decadal-programme2/dagstuhl-workshops

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

on a set of use cases that do not represent “domains” but “communities of practice”. The use cases identified during the 2021 Dagstuhl workshop are summarised below and indicate that DCAT offers potential for supporting findability for different communities of practice and shows how DDI-CDI might facilitate interoperability.

The idea of a Core Interoperability Framework is, in fact, the point at which all the other use cases of the workshop intersect. It looks at each FAIR principle and analyses which set of cross-domain standards and models would better support it. It is very much a work in progress.

- **Findability**: standards like DCAT (including DCAT-AP) and *schema.org*, as well as DOIs for persistent identification.
- **Accessibility**: Open Digital Rights Language (ODRL, v. 2.2); and Data Privacy Vocabulary (DPV, under development, v. 0.3).
- **Interoperability and Reuse**: As labour-intensive, automation might be more efficient.
    - **Structural Metadata**: DDI-CDI and other more limited standards (W3C CSV on the Web; W3C Data Cube Vocabulary/SDMX…).
    - **Semantics**: OWL, RDF-Schema or Simple Knowledge Organization System - SKOS (XKOS for statistical classifications). Vocabularies and classifications/thesauri are useful for attaching semantics to the structures of data.
    - **Context and Fully Described Observations of Interest**: W3C PROV Ontology (PROV-O) and some data-specific profiles (e.g., PROV-ONE, etc.). Also, clusters of variables help with specific measurements or observations. There are standards and models related to “observable properties” (RDA’s I-ADOPT working group).

4.3 Use cases for interoperability

4.3.1 Helmholtz Metadata Collaboration (HMC) Use Case[37]

This use case (still exploratory) includes earth scientists in Helmholtz centres that intend to apply FAIR data sharing within their community. One of the complexities of this use case is the number of institutions working together, many of which already maintain infrastructures (such as Pangaea or ICAT).

Initially, their approach was based on *schema.org* and implemented using JSON-LD. Now, to cover integration and reuse, they are looking at a more granular data description with DDI-CDI. Precisely, the proposal is to describe the data set in OWL (based on the CDI model). This becomes a reference point for further work on the FDO description. The Kernel of Information might be DCAT or *schema.org*, with CDI structural metadata and a PROV store.

4.3.2 European Social Survey (ESS) Use Case[38]

This use case focuses on improving the integration of ESS data with data on environment, health, and other contexts from European sources. It looks at the ESS Multi-Level application, and how the combination of DDI-CDI with existing metadata standards (particularly SDTL) and applications. The

[37] https://helmholtz-metadaten.de/en
[38] https://www.europeansocialsurvey.org/about

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

method is based on leveraging the datum-oriented approach within DDI-CDI, and the process components, to try to increase automation and efficiency.

Among all the limitations that are mentioned in this use case, we find some similarities towards the testimonials in this deliverable. For instance, difficulty with time and geospatial variables in, for example, census fields (also mentioned by B2FIND). In addition, the versioning issue, in this case, is faced with the DDI3.3 variable cascade (instance variable, represented variable and conceptual variable) but it presents limitations towards tracking changes at a cell level.

4.3.3 Smart Energy Research Laboratory (SERL) Use Case[39]

This use case analyses how to combine the granular management and dissemination of data with the automation, in order to improve manual integration across domain boundaries, at the dataset level. Also, it discusses data services needs to support variable subsetting (benefits for data reuse across domains). The main approach consists of common object descriptions in schema.org or DCAT, including a link to PROV information as part of a "catalogue" record. Geo/Conceptual coverage for a single dataset can be served by DDI. Regarding datasets with very different structures, the structure can be harmonised with CDI to allow subsetting. Machine Learning techniques could be applied in the future.

The SERL use case also points out the need of a metadata improvement, which implies the enrichment of descriptive, administrative, structural, process and variable metadata, as well as permissions and rights (a similar claim is made by CESSDA in this deliverable's interviews).

4.3.4 Interstat Census Example / NGSI-CDI-SDMX-LD Use Case[40]

The InterStat project is looking at incorporating official statistical data and other types of research data into general public information produced by European governmental bodies. The focus is on bridging across the set of domain specific models and standards to support the information channels being employed by the EC. This spans SDMX, DataCube, the NGSI-LD domains, DDI-CDI, SOSA-SSN, and potentially other standards.

In this use case, the features and extensions of DCAT are explored, including some recommendations for DCAT (structuredBy and documentedBy and conformsTo properties) and a proposal for inclusion in DCAT v3.

Also, ENVRI-FAIR[41] and EOSCLife[42] are new proposed use cases to be explored within this workshop.

[39] https://serl.ac.uk
[40] https://www.eng.it/en/case-studies/interstat-open-statistical-data-interoperability-framework
[41] https://envri.eu/about-envri-fair/
[42] https://www.eosc-life.eu/

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

5. Conclusions and ongoing work

Based on the findings of this small scale pilot and the preceding work presented in deliverable D3.6 "Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake", it is becoming clear that there is a potential role for using DCAT 2 to support the aggregation and findability of metadata catalogues. It is clear that DCAT 2 on its own will not solve the larger problem of interoperability at the dataset level but there are promising signs that a combined approach of using DCAT 2 at the generic level and DDI-CDI at the domain specific level could provide a solution for supporting both the findability and interoperability of heterogeneous research data.

From our interviews with representatives from two of the thematic cluster projects (SSHOC and PaNOSC) it seems that adoption of DCAT 2 is feasible as it maps well to SSH and Photon/Neutron community standards. The findings of the Dagstuhl workshop show that other disciplines also see potential value in using DCAT. However, there are legacy issues relating to metadata quality and shared access to community owned, canonical metadata mappings that need to be addressed to facilitate progress. FAIRsFAIR carried out a review of available metadata mappings as part of the pilot and there are several that are currently being developed or being maintained at the grass roots level. Given the time and effort required to carry out metadata mappings and the potential for subjectivity and differing views on the relationships between fields, we believe there is a need to better coordinate activity on developing and maintaining these mappings to avoid duplication of effort, result in more authoritative mappings and, most importantly, to ensure that these resources can be found and reused by those who would benefit from them. To this end, FAIRsFAIR will seek to continue our dialogue with relevant actors to discuss possible ways to progress a more cohesive approach to metadata mappings and to consider how these valuable resources might be made more visible, citable, and reusable (e.g., making them available in registries such as FAIRsharing[43]).

From the perspective of the two service providers (B2FIND and OpenAIRE), the main obstacle hindering adoption is the current lack of demand for harvesting DCAT 2 from the repository communities they serve. Implementing DCAT 2 harvesting capabilities could be prioritised if the demand from the community was strong. As such, there is more work needed to ensure that repository communities are aware of DCAT 2 and its potential value for supporting metadata catalogue integration. To this end, FAIRsFAIR will explore running a Repository Support Webinar[44] with representatives of DCAT to help increase awareness and to share the forthcoming revisions being undertaken in DCAT version 3. By helping to raise awareness among the community, we may see demand for DCAT harvesting increase and help aggregator services to make the case for allocating resources to the development of such functionality.

The efforts described above will be progressed over the final months of the FAIRsFAIR project which will end in February 2022. From the outset of the metadata catalogue integration activity, FAIRsFAIR has placed an emphasis on being inclusive and has sought active involvement from domain representatives and service providers. This collaborative approach will continue to be our focus as

[43] https://fairsharing.org/
[44] https://fairsfair.eu/events/webinars/repository-support-webinars

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

we progress and will hopefully mean that community engagement on this topic will continue beyond the life of the FAIRsFAIR project.

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

Bibliography

Becker, A., Loeden, A., Manghi, P., Principe, P. & Schirrwagen, J. (2018). OpenAIRE Content Acquisition Policies. https://doi.org/10.5281/zenodo.1446407

EOSC Executive Board (2021). EOSC Interoperability Framework: Report from the EOSC Executive Board Working Groups (WG) FAIR and Architecture. https://doi.org/10.2777/620649

Haynes, D., 2018. Metadata for Information Management and Retrieval: Understanding metadata and its use, 1st ed. Facet. https://doi.org/10.29085/9781783302161

Goble, C. & Juty, N. (2020). Analysis of existing research data cataloguing efforts towards integrated discovery. EOSC Enhance deliverable D4.3. https://repository.eosc-portal.eu/index.php/s/gFWTQK26FPw5No9

Gotz, A., Petzold, A., Asmi, A., Blomberg, N., Lamanna, G., & Dekker, R. (2020). ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC. https://doi.org/10.5281/zenodo.3675081

Gregory, K., Cousijn, H. & Groth, P. (2019). Understanding data search as a socio-technical practice. Journal of Information Science 46(4). https://doi.org/10.1177/0165551519837182

Grootveld, M., Hodson, S., Pittonet, S., Davidson, J. & Dillo, I. (2021). Report 3 of the Synchronisation Force. FAIRsFAIR deliverable D5.6. https://doi.org/10.5281/zenodo.5336658

Jones, S., Dimper, R., Dillo, I., Hanahoe, H. & Kurapati, S. (2021). Making the European Open Science Cloud (EOSC) work: where to go from here? https://doi.org/10.5281/zenodo.5607692

Lambert, S. (2018). First Annual Report on PID Commons and Sustainability. FREYA deliverable D6.1. https://doi.org/10.5281/zenodo.3250701

Mendez, E., Hernandez, T., Whyte, A., & Davidson, J. (2020). Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake (1.0 DRAFT). FAIRsFAIR deliverable. https://doi.org/10.5281/zenodo.4134788

Tompkins, V.T., Honick, B.J., Polley, K.L., Qin, J. (2021). MetaFAIR : A Metadata Application Profile for Managing Research Data. Proc. Assoc. Inf. Sci. Technol. 58, 337–345. https://doi.org/10.1002/pra2.461

Wimalaratne, S. & Ulrich, R. (2021). M4.8 Introduce additional components and practices to metadata schema and align them with FAIR data practices (1.0). https://doi.org/10.5281/zenodo.5473107

Zeng, M. L., & Qin, J. (2016). *Metadata* (2nd edition). Neal-Schuman.

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

Annex 1: Interview questionnaires

1. For interviews with (meta)data providers

FAIRsFAIR interviews on metadata catalogue integration September 2021

Questions to guide the interviews

1.	How do you currently make your metadata harvestable by aggregators? (or do you have plans to do that or to enhance it?) What metadata do you currently expose?
2.	Do the classes of DCAT 2 (*Catalog*, *CatalogRecord*, *Dataset*, *Distribution* etc.) map well to your handling of datasets? Note: A diagrammatic overview of DCAT 2 will be available at the interview, or see Figure 1 on the page https://www.w3.org/TR/vocab-dcat-2/
3.	What **domain-agnostic** aspects of your current metadata standards are most important for discoverability of your data? How well do they fit with DCAT 2? Are there any gaps?
4.	What **domain-specific** aspects are most important for discoverability of your data?
5.	How could you envisage the interface between a domain-agnostic standard such as DCAT 2 and domain-specific aspects for enhancing discoverability of datasets?
6.	Are there any barriers or challenges that would impede you using DCAT2?
7.	Do you have any views on the proposed changes in DCAT 3 (principally relating to versioning and dataset series)? Note: Major changes are the extension of DCAT 2 with terms and guidelines on *versioning* and *dataset series*.
8.	Are there any other issues around metadata catalogue integration that you would like to raise?

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

2. For interviews with metadata aggregator services

FAIRsFAIR interviews on metadata catalogue integration October 2021

Questions to guide the interviews

1.	How do you currently harvest/aggregate metadata? Any plans to extend that?
2.	Do the classes of DCAT 2 (*Catalog*, *CatalogRecord*, *Dataset*, *Distribution* etc.) map well to your handling of harvested/aggregated metadata? Note: A diagrammatic overview of DCAT 2 will be available at the interview, or see Figure 1 on the page https://www.w3.org/TR/vocab-dcat-2/
3.	What aspects of metadata do you think are most important for discoverability?
4.	(Question intentionally omitted)
5.	How could you envisage the interface between a domain-agnostic standard such as DCAT 2 and domain-specific aspects?
6.	Are there any barriers or challenges that would impede you using DCAT2?
7.	Do you have any views on the proposed changes in DCAT 3? Note: Major changes are the extension of DCAT 2 with terms and guidelines on *versioning* and *dataset series*.
8.	Are there any other issues around metadata catalogue integration that you would like to raise?

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

Annex 2: Workshop on Metadata Catalogue Integration for Interdisciplinary Research (March 2021)

List of participants

Name	Affiliation and project/infrastructure/service represented
Isabel Bernal	CSIC, DataCite Metadata Working Group, COAR Vocabularies, EOSC SYNERGY)
Ricarda Braukmann	DANS, FAIRsFAIR
Joy Davidson	DCC, FAIRsFAIR
Paul Finglas	FNS-Cloud
Anna-Lena Flügel	DKRZ, B2FIND
Alejandra Gonzalez-Beltran	STFC, W3C Dataset Exchange WG, ExPaNDS project
Arofan Gregory	CODATA
Simon Hodson	CODATA
Nick Juty	University of Manchester, ELIXIR; EOSCLife, FAIRplus, EOSC Enhance
Mari Kleemola	CESSDA/FSD, SSHOC project
Barbara Koroušić Seljak	FNS-Cloud
Keith Jeffery	Keith G Jeffery Consultants, EPOS, ENVRI
Simon Lambert	STFC, FAIRsFAIR
Mireille Louys	Observatoire Strasbourg, Astronomical Virtual Observatory project, ESCAPE project
Paolo Manghi	OpenAIRE
Claudia Martens	DKRZ, B2FIND
Agnieszka Matuszczak	FNS-Cloud
Eva Méndez	UC3M, FAIRsFAIR
Jessica Parland-von Essen	CSC, FAIRsFAIR

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

Pedro Principe	University of Minho, FAIRsFAIR, OpenAIRE
Marina Sánchez Moreno	UC3M, FAIRsFAIR
Jutta Schnabel	ESCAPE project
Carsten Thiel	CESSDA
Maria Traka	Quadram Institute Bioscience, FNS-Cloud
Joachim Wackerow	GESIS, DDI Alliance
Danielle Welter	University of Luxembourg, ELIXIR, FAIRplus
Angus Whyte	DCC, FAIRsFAIR
Peter Winstanley	Semantic Arts; W3C Dataset Exchange WG

Programme of the workshop

15m Introduction: reminder of the metadata catalogue integration pilot (Simon Lambert)
20m The background: previous workshops, relevant existing work, the wider context (Eva Méndez)
10m Applying DDI-CDI (Data Documentation Initiative-Cross Domain Integration) to the EOSC (Arofan Gregory)
15m General questions and discussion
Short break
60m How to participate in the pilot and what the participants should do
Short break
30m Use cases and development of an assessment framework
30m Planning for the immediate future: overall timetable, next workshops, takeaway actions

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