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

D2.1 REPORT ON FAIR REQUIREMENTS FOR PERSISTENCE AND INTEROPERABILITY 2019

Work Package	WP2 - FAIR practices: semantics, interoperability and services
Lead Author (Org)	Heikki Lehväslaiho (CSC)
Contributing Author(s) (Org)	Jessica Parland-von Essen (CSC), Claudia Behnke (SURFsara), Leah Riungu-Kalliosaari (CSC), Heidi Laine (CSC), Yann Le Franc (e-SDF), Christine Staiger (DTL)
Due Date	30.11.2019
Date	23.11.2019
Version	1.0
DOI	https://doi.org/10.5281/zenodo.3557381

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)

Abstract

This document is the first iteration of three annual reports on the state of FAIR in European scientific data by the FAIRsFAIR project. The interpretation of the FAIR data principles and their implications for services are now under intense scrutiny across Europe with multiple possible outcomes. The report is based on studies of public information, especially EOSC infrastructure efforts, and on limited surveying and interviews. The focus has been on understanding the usage of persistent identifiers and semantic interoperability. This study highlights the rapidity of change in technical solutions and wide variation across scientific domains in the uptake. More efforts are needed to guide researchers in best practices.

Versioning and contribution history

Version	Date	Authors	Notes
0.9	30.10.2019	All contributors	Draft for internal review
1.0	23.11.2019	Heikki Lehväslaiho, Heidi Laine and Jessica Parland-von Essen	Content ready

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

Summary

This report is the first of three of a kind to be produced by the FAIRsFAIR project. This deliverable reviews and documents commonalities and possible gaps regarding semantic interoperability, and the use of metadata and persistent identifiers across infrastructures. Since many landscaping, specification and "FAIRification" activities are ongoing in the EOSC projects and elsewhere, much new information will be added to the later versions. The authors hope to get feedback to enrich and adjust the observations and conclusions made in this document.

FAIR Digital Objects are central to the realisation of FAIR data principles. These objects need to be accompanied by Persistent Identifiers (PIDs) and rich metadata as they sit in a wider FAIR ecosystem comprising of services and infrastructures for FAIR, including identifiers, standards and repositories. The details of the FAIR principles for data, the implementation and implications for services are neither defined nor settled yet. The first suggestions for a more specific definition of a FAIR Digital Object has only recently been presented and will be further tested within the FAIRsFAIR project. Implications of the FAIR data principles for services, repositories and software are being investigated in other FAIRsFAIR tasks. Thus, this report focuses on semantic interoperability as it is a prerequisite for linking and finding data, as well as on the identifiers, which can offer persistence but also need context sensitive solutions. We use the term semantic artefact to overcome the terminological diversity that ironically is a challenge in discussions on this important element of the architecture we need in order to enable semantic interoperability within a FAIR Ecosystem.

Development and implementation of the FAIR data principles should be driven by researcher needs to achieve wide penetration and the potentially significant benefits of FAIR data. The differences within research domains are often bigger than between them. Enforcing standards comes with the risk of making gaps grow between mature and emerging research domains. Community adoption and trust are decisive factors. Enabling services for publishing crosswalks, mappings and semantic application profiles are needed. All these should be registered and published in machine readable formats. A challenge with PID and data type registries is having them to promote reuse of data rather than bulk creation of PIDs. To support interoperability, they should be considered semantic artefacts, curated and reused. The aim should be born-FAIR data, which requires integrated and user friendly solutions throughout the research process and data lifecycle.

By publishing application profiles, preferably in a common registry and in a machine readable format, reuse of semantic artefacts can be promoted, thereby enabling interoperability. Also curated registries like the EOSC Hub, FAIRsharing and re3data.org are important resources for enabling implementation of the FAIR data principles.

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

1. Introduction **6**
1.1. Background and scope 6
1.2. Methods 7
1.2.1. Desk research 8
1.2.2. Survey 9
1.2.3. Interviews 11
2. The elements of FAIRness **12**
2.1. The FAIR technologies and methods 13
2.1.1. Semantic interoperability 15
2.1.2. Semantic artefacts 17
2.1.2.1. Interoperability for semantic artefacts 19
2.1.2.2. CASE: OBO Foundry recommendation for ontologies 21
2.1.3. PIDs and PID services for research data 23
2.1.3.1. CASE: Recommendations about Persistent Identifiers 25
2.1.3.2. CASE: DiSSCo 26
2.2. FAIR in the context of the Data Life cycle 29
2.2.1. Data Repositories 30
2.2.2. Evolving datasets and data citation 31
2.2.2.1. CASE: Evolving dataset citation 32
3. The current status of FAIR data at a glance **34**
3.1. International efforts to promote the FAIR principles 34
3.1.1. CASE: FAIR according to Turning FAIR into Reality 34
3.1.2. EOSC 35
3.1.3. FORCE11 36
3.1.4. GO FAIR 37
3.1.5. FAIRsharing 37
3.1.6. DataCite 38
3.1.7. re3data.org 38
3.1.8. FREYA and Open Citations 39
3.1.9. Research Data Alliance 39
3.1.9.1. Interest groups 40
3.1.9.2. Working groups 40
3.1.9.3. Other RDA groups 41
3.1.9.3.1. CASE: RDA FAIR Data Maturity Model Working Group 41
3.1.10. Other groups and stakeholders 43

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

3.1.10.1. CASE: List of technologies monitored by COAR 44
3.2. The landscape of digital research infrastructures 45
3.2.1. Energy 46
3.2.1.1. Metadata 47
3.2.1.2. Semantic interoperability and artefacts 47
3.2.1.3. Identifiers 47
3.2.1.4. CASE: Energy research 47
3.2.2. Environment 48
3.2.2.1. Metadata 49
3.2.2.2. Semantic interoperability and artefacts 49
3.2.2.3. Identifiers 50
3.2.2.4. CASE: AgroPortal 51
3.2.3. Health & Food 51
3.2.3.1. Metadata 55
3.2.3.2. Identifiers 55
3.2.4. Physical Sciences & Engineering 55
3.2.4.1. Metadata 58
3.2.4.2. Semantic interoperability and artefacts 58
3.2.4.3. Identifiers 59
3.2.4.4. RDA CHEMISTRY IG 59
3.2.5. Social & Cultural Innovation 60
3.2.5.1. Metadata standards 61
3.2.5.2. Semantic interoperability and artefacts 62
3.2.5.3. Identifiers 63
3.2.5.4. CASE: CESSDA 63
3.2.5.5. CASE: Europeana 65
3.2.6. Data, Computing and Digital Research Infrastructures 66

4. Conclusions 66

5. Bibliography 69

6. Appendix A. Acronyms and abbreviations 72

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

Connecting electronic data stores together by the Internet has seemingly made everything possible. In practice, it has highlighted all the existing problems in research data management, especially in interoperability of separate systems. Each scientific domain and subdomain develops its own language to describe its subjects. Nomenclatures differ in fundamental ways even in closely similar topics. The growing amounts of data have created the FAIR principles for research data (Wilkinson et al., 2016). A perfect data management would create digital research that is reproducible and resources that are reusable. In reality data is often hard to discover and difficult to reuse, which causes harm both to quality and efficiency in research.

At the advent of the Human Genome Project, it was painfully obvious that the combined knowledge on the function of genes that were stored in several model organism databases was not semantically interoperable. While their combined knowledge was needed to annotate the upcoming complete collection of human genes, the nomenclature, the system of giving names, within each model organism was too different and terminology used to describe their function was too confusing.

The solution came through the insight of Michael Ashburner, professor of genetics in Cambridge, UK, who saw that gene function description in scientific parlance needs to be separated into three distinct but mutually supporting areas: cellular component, molecular function and biological process. Also, these were not described in flat or unstructured lists that had been common, but in well defined, hierarchical structures (directed acyclic graphs) to form ontologies.

The creation of the Gene Ontology from the combined knowledge of fly, yeast and mouse genome databases was the first practical, scientific implementation of the concept of a machine readable formally defined ontologies and proved immensely powerful (Ashburner et al., 2000). It opened the world of semantic interoperability of data to wider application. We are currently trying to understand its implications for practical data management problems.

1.1. Background and scope

While the past decades of discussion around scientific data management at least on a policy level was on around open data, the seminal 2016 paper expressed data management and re-use problems in richer terms by dividing them into four main principles: findable, accessible, interoperable and reusable (FAIR)(Wilkinson et al., 2016). This conceptual innovation received wide approval. The ongoing work on further defining, measuring and applying these FAIR principles to day-to-day workings of scientific knowledge sharing and dissemination form the landscape this report tries to illuminate.

As part of the EOSC projects ecosystem, the FAIRsFAIR - Fostering Fair Data Practices in Europe - project aims to supply practical solutions for the use of the FAIR data principles throughout the research data lifecycle. The FAIRsFAIR project lays emphasis on fostering FAIR data culture and the

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

uptake of good practices in making data FAIR, but there are still many discussions to have on what the implementation of the FAIR data principles actually means, for instance for services and the digital research infrastructures as an ecosystem. It is important to look at not only data management practices but also to find solutions that are resilient over time.

This report is the first in a series of three versions that will be progressively reviewing the state of the art in the technical implementation of the FAIR principles. This report focuses on solutions for semantic interoperability and on persistent identifiers as they are important building blocks of a FAIR ecosystem and framework. We review the implementation of semantic interoperability and persistent identifiers in projects and landmarks listed by the European Strategy Forum on Research Infrastructures (ESFRI[1]). The issues addressed include commonalities and gaps among the ESFRI projects regarding standards for and implementation of semantic interoperability, vocabularies and ontologies, metadata, and persistent identifiers. Still, a large amount of the work in this field is done in projects outside the European Open Science Cloud (EOSC) and in cooperation with global partners and communities. Hence, we take a broader perspective and have included for instance much of the important work done in the Research Data Alliance (RDA).

The two subsequent reports will broaden the current scope. There are many relevant tasks and projects starting within other EOSC projects and also the work on global standards and good practice is progressing. The outputs of this work will be included in the coming reports. Some parts will update and revisit the findings of this report, and other parts will open new lines of investigation, such as broadening the number of reviewed projects and communities, and examining how F, A, I, & R are being measured within European member states and research communities. However, our first goal has been to paint a picture of the landscape as a whole.

Which implications the FAIR principles have on delivering and developing research data services and infrastructures is not settled yet. “FAIRness” regarding semantics, services, software and repositories is being discussed and formulated for example in the other tasks of the same work package that has produced this report. Also, important deliberation around implementation and evaluation is done within the RDA maturity group, which will be presented below.

1.2. Methods

As the definition of the FAIR principles for other parts of the framework and ecosystem than data is not agreed upon, this report describes the landscape of semantic interoperability and persistence of research data management solutions. The field is vast and diverse, so focus is kept on research data and its formats and life cycle rather than research information. Information was gathered through three different efforts (Figure 1):

1. Desk research
2. Survey data
3. Interviews with focus digital research infrastructures

[1] European Strategy Forum on Research Infrastructures. [web page] ESFRI. [cited 9.10.2019] Available from: https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/esfri_en

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

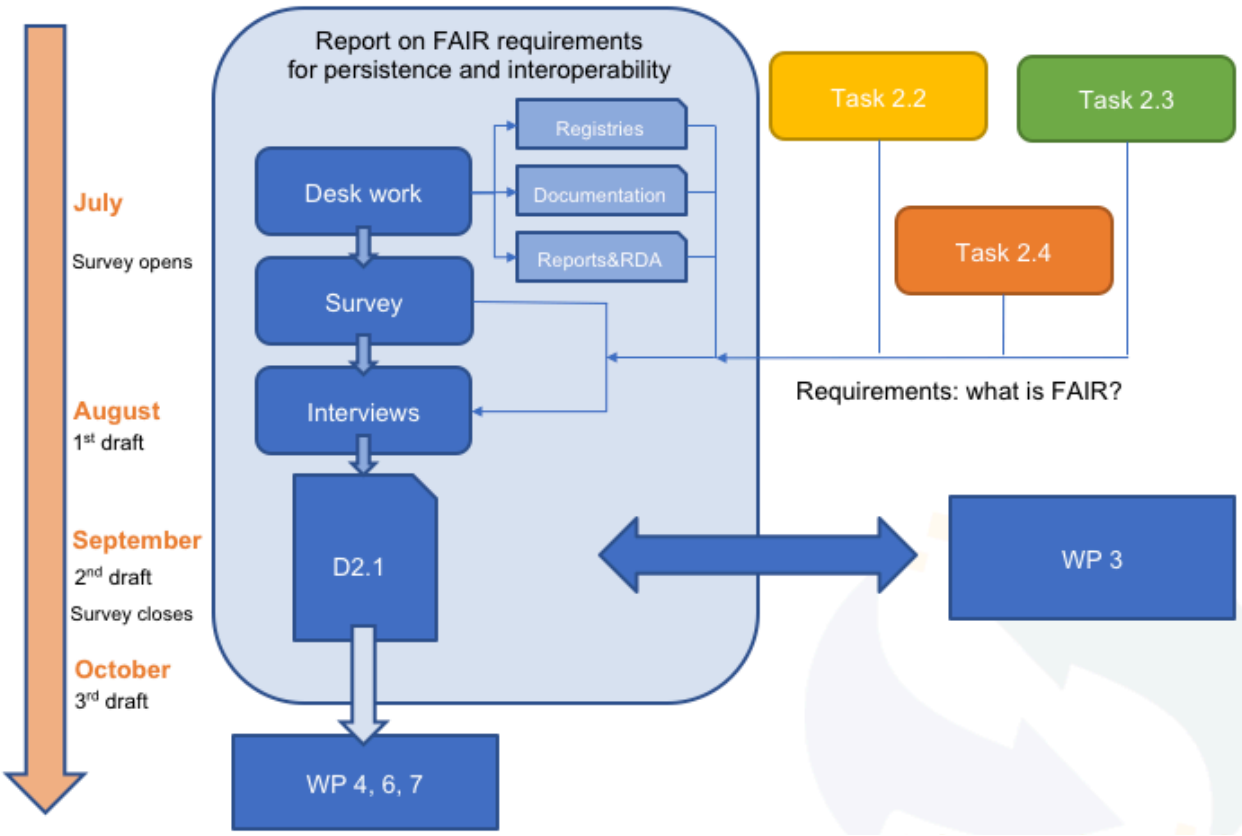

Figure 1. The process for creating this report.

1.2.1. Desk research

Research infrastructures are changing rapidly due to the rise of enormous amounts of data. There are several simultaneous efforts to tackle difficulties with research data that is hard to find and not readily (re)used. Often the difficulties are related to insufficient metadata (Chataigner and Nowak, 2018; Gregory et al., 2019; van Raaij, 2018). That problem however stems more from the data management practices, that are covered in other reports, especially the report D3.2 FAIR Practice Analysis, also published by FAIRsFAIR in parallel with this report. We have excluded the policy level issues as well as questions about support services and training for researchers, since those questions also will be handled in other tasks and projects. For our desk research we chose three different types of sources:

1. Through previous landscaping efforts and overviews, we located existing infrastructures that have relevant data infrastructure
2. We looked at FAIRsharing and the documentation of metadata and persistent identifiers (PID)
3. We examined RDA groups and their outputs, and stakeholders we identified through them (CODATA, etc)

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 wanted to find evidence and examples of solutions and methods to improve semantic interoperability by methods or technologies for describing and publishing data. We looked for the following as we went through the material:

1. Descriptions and definitions of FAIR in technical contexts
2. Documentation about metadata, application programming interfaces (API) and PIDs for research data (in use or plans)
3. Technical implications to be considered to reach FAIR data
    a. Standards
    b. Semantic interoperability
    c. Vocabularies and ontologies
    d. Metadata
    e. Persistent identifiers (PIDs)
4. Expressed problems and uncertainties in implementing FAIR

1.2.2. Survey

To complement and validate information from the desk research, we created a survey that was aimed at data managers and data support experts. We aimed to collect information about tools and services we might have missed, but also wanted reflections on the thinking around identifiers and ontologies and other semantic artefacts. The information will also help in preparing the workshops on semantics and interoperability that are forthcoming within the project, as well as the work on software and services. The survey covered questions about metadata, use of persistent identifiers, use of semantic artefacts and handling of research software. The data and questions are published in Zenodo (Lehväslaiho et al., 2019). The survey was conducted as a joint effort with WP3 and it was disseminated on the fairsfair.eu web pages, social media channels and email lists.

We received 66 answers during the period the survey was open, that is between 15 July to 2 October 2019. The roles of the people submitting answers were the following (it was possible to choose several options):

Research support staff 28
Repository staff 19
Research infrastructure operator 17
Researcher 22
Policy maker 5
Other 5

Research support staff and repository staff was the most common combination of roles. The largest groups of people with just one role were researchers (12) and research support staff (13). Other roles mentioned were data manager, data steward, stakeholder, technical coordinator and software development manager.

Many responses covered several research domains. A minority of the responses (28) were not related to any specific infrastructure. Infrastructures mentioned were:

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

ACTRIS ADC AnaEE (2) BBMRI ERIC CESSDA ERIC (5) CLARIN (4) DARIAH (3) DiSSCo (3) EISCAT_3D ELIXIR (3) eLTER (2)	EMBRC ERIC EMODnet EMPHASIS (2) EPOS ESS ERIC EU-SOLARIS EURO-ARGO ERIC (2) European XFEL FAIR (3) Go FAIR Initiative IAGOS (2)	INSTRUCT ERIC (2) IODE IS-ENES (2) LifeWatch ERIC ODP OpenAIRE PRACE (3) SCADM SeaDataNet/SeaDataCloud (2) SHARE ERIC SKA SOOS

Table 1: Infrastructures represented in the survey (number of times mentioned)

The geographic coverage was spread out as follows: Germany (12), Netherlands (10), France (8), UK (7), Finland (6), other European countries (24), other countries (8).

Number of researchers in organization	Number of responses (N=64)
< 100	16
100 - 500	13
500 - 1000	7
1000 <	28

Table 2: Responses across the organizations

It is important to understand that the data in this survey is not viable for any quantitative analysis. The information about the infrastructures and fields of the different scientific domains will be discussed separately in the domain context below. Figure 2 shows how many times the respective domains are mentioned.

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

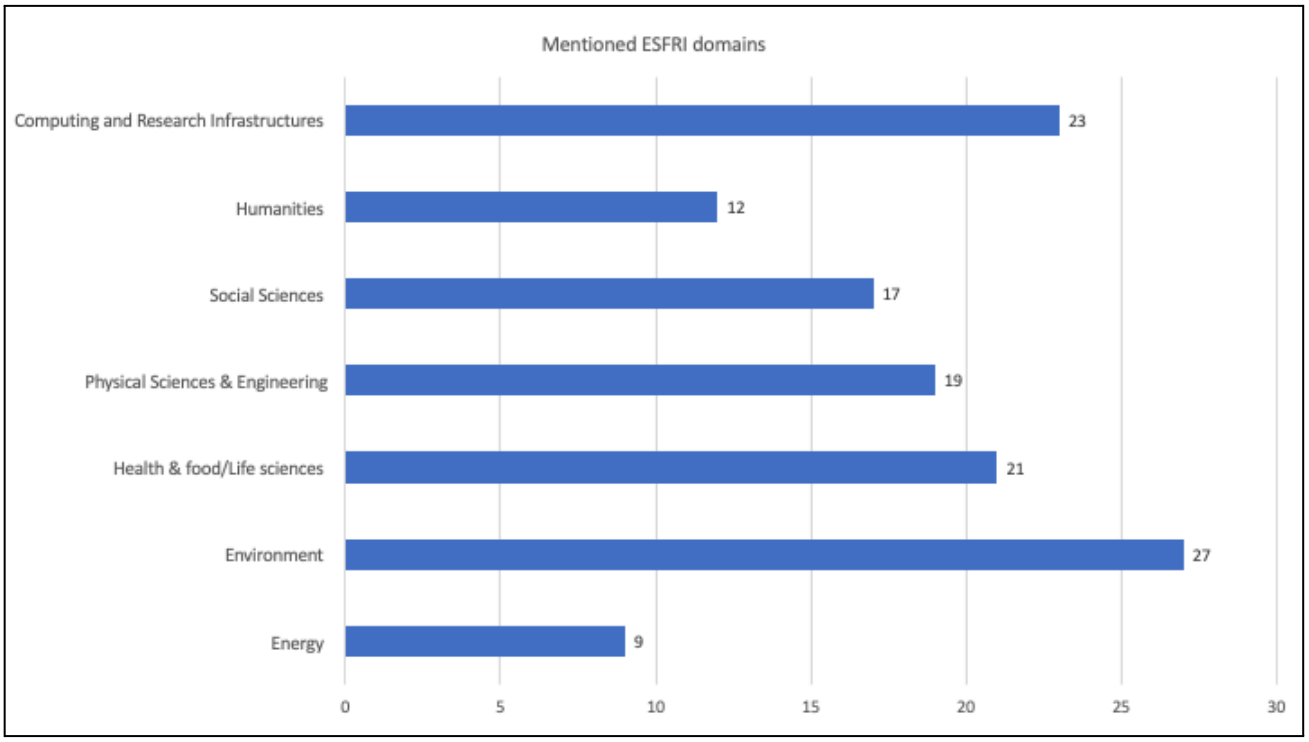

Figure 2: Infrastructures and scientific domains

1.2.3. Interviews

We conducted five semi-structured expert interviews during September 2019. The aim was to collect views from agents that work with interoperability and services that are domain agnostic or domain independent and generic. We also made one interview for one of the case studies (DiSSCo). The notes were deleted and the final text from this report was checked by the interviewees. The experts came from the following institutions or projects: DataCite, Deutsches Klimarechenzentrum (DKRZ), DiSSCo, FAIRsharing and Figshare. Their opinions must of course not be considered as representing that of their affiliation, but are stated as opinions of experts with a high level of insight and from vantage points in respect to questions of technical solutions for interoperability.

Our aim was to get a better understanding of

- Practical implementations of semantic interoperability across infrastructures
- What are seen as the most critical factors for success in FAIR and semantic interoperability
- What are the most serious omissions in currently available tools and specifications

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. The elements of FAIRness

For a data resource to be considered a FAIR digital object, it needs to be accompanied by persistent identifier(s) and metadata rich enough to enable it to be unambiguously understood, used and cited, following metadata standards and vocabularies adopted by the related research community. In addition, the data needs to be represented in common, and ideally open, format[2].

For metadata there are some recommendations produced within the Metadata 2020 project.[3]A metadata standard, to help data to be FAIR should according to them:

- use of formal, accessible, shared, and broadly applicable schema
- support the production of a rich description of the data
- support data citation, i.e. include necessary information elements to create an actionable reference pointing to the data resource, that is both human and machine readable
- respond to disciplinary needs; a metadata standard needs to suit the data, not vice versa
- be commonly used and known among the relevant community
- be open, i.e. it should be freely available

In terms of identifiers to reach FAIRness it requires:

- being persistent i.e representing a resource even when the resource changes, gets updated or is no longer available online
- uniqueness
- actionability

Vocabularies and ontologies need to provide commonly agreed-upon terminology and concepts serving as a basis for implementing FAIR capabilities. There is another task (T2.2) dedicated to this topic within the FAIRsFAIR project. To be considered FAIR, a technical repository solution should provide an API with the capabilities to support FAIR data principles. Another work task (T2.3) is looking into helping repositories to achieve better degrees of FAIRness. A formal technical definition of FAIR Digital Objects is still missing.[4] One model has been presented in 2018 within the RDA GEDE group (Figure 3). A digital object may represent data, software or other research resources.

[2]TeD-T, the Term Definition Tool of the Data Foundation and Terminology Interest Group (DFT IG) of the Research Data Alliance (RDA). Vide FAIR Digital Objects. [web page] [Cited on 3.10.2019] Was available from: https://smw-rda.esc.rzg.mpg.de/index.php?title=FAIR_Digital_Objects

[3] Metadata 2020. [web page] Available from: http://www.metadata2020.org/

[4] This term presented in the report Turning FAIR into reality has been adopted by the EOSC Secretariat. See also [webpages] Meerman B, FAIR DIGITAL OBJECTS driving worldwide interoperability. 2019. [cited 22.11.209] Available from: https://www.eoscsecretariat.eu/eosc-liaison-platform/post/fair-digital-objects-driving-worldwide-interoperability%C2

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

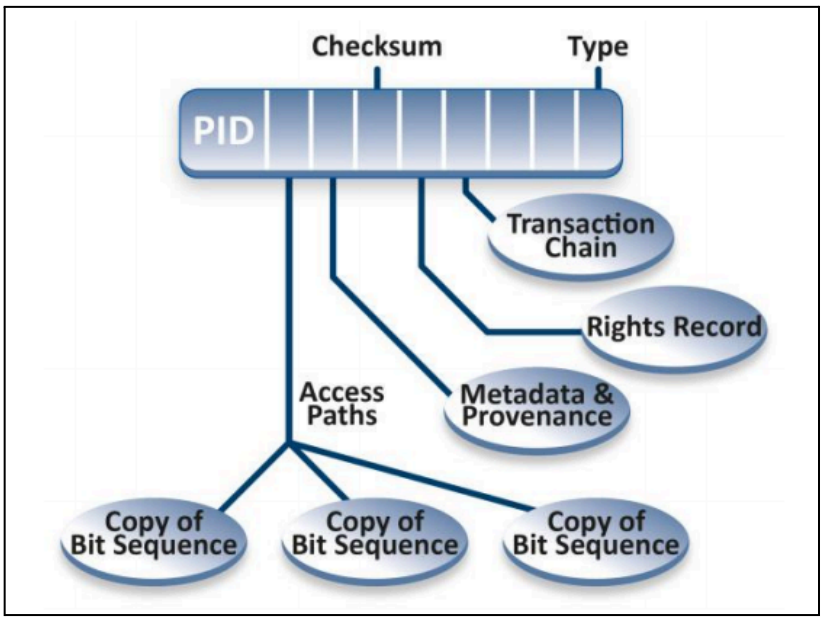

Figure 3. The "FAIR Objects" that has been understood as a synonym for "FAIR Digital Objects."[5]

2.1. The FAIR technologies and methods

The FAIR principles for data were initially described as a concept and have not been well defined at implementation level. FAIRness should be seen as a continuum where increased alignment is a continuous process towards an unattainable ideal. Any action that increases FAIRness of data is an improvement.

While most common approaches to FAIR data rely on Linked Data and related technologies and W3C standards (e.g. OWL, SKOS, RDF, SPARQL) it is important to remember that "FAIR is not equal to RDF, Linked Data, or the Semantic Web", but that "resources that wish to maximally fulfil the FAIR guidelines must utilise a widely-accepted machine-readable frameworks for data and knowledge representation and exchange. While there are only a handful of standards and frameworks that could, today, fulfil this requirement, other potentially more powerful approaches may appear in the future" (Wittenburg et al., 2018).

While tools specifically developed for Semantic Web offer the most comprehensive approach to data management for both machine and human readability, they contain concepts and approaches that are radically different from legacy relational database approach that has been commonly and widely used in sciences. REpresentational State Transfer[6] (REST) architecture for interchange and JSON-LD[7] for expression of metadata statements can help to bridge this gap.

[5] Wittenburg P, Strawn G, Mons B, Bonino L, Schultes E. Digital Objects as Drivers towards Convergence in Data Infrastructures. [presentation] RDA. 2018. Available at: https://www.rd-alliance.org/sites/default/files/Digital%20Objects%20as%20Drivers%20towards%20Convergence%20in%20Data.pdf

[6] REST [web page] Available from: https://en.wikipedia.org/wiki/Representational_state_transfer

[7] JSON-LD [web page] Available from: https://json-ld.org/

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

The simple, underlying structure of the Resource Description Framework (RDF) is a semantic triple that contains three entities giving statements in the form of subject-predicate-object. All entities can be represented by unique URIs. Predicates are verbs in this statement describing the kind of relationship. Multiple triples form a network, a graph that describes a dataset.

Every entity has a unique identifier that is stored in a semantic collection. These collections are called various names that include glossary, controlled vocabulary, controlled vocabulary, thesaurus, data models. In this report, we use the term semantic artefact[8] to cover them all.

The philosophy of RDF assumes a flexible, multi-source and multi-consumer world. RDF was created to be flexible to the extent that there are no absolutely correct ways of expressing data relationships. Most of the complexities of RDF arise from this fact. A good example of this flexibility is validation of RDF. Numerous technologies within the framework allow for limited ways of defining restrictions to triples and their relationships (RDF itself, SKOS, Data Cube Vocabulary, R2RML, RDFS, OWL, ICV, SPIN/SPARQL, ShEx, SHACL).

The Shapes Constraint Language[9], SHACL, tries to make it easier than other RDF validation options to write simple statements about data. Using its own terminology, it allows the writing of "shapes graphs" against "data graphs". In practice, SHACL makes it possible to describe what properties and relationships nodes in the graph must have and or must not have, use them to filter the graph, and raise an error when these conditions are not met.

Many of the recommendations and insights for using RDF come from Research Data Alliance[10] (RDA) Working Groups and Interest Groups[11]. The RDA Data Fabric Interest Group (DFIG) looks at the data creation and consumption cycle to identify opportunities to optimize the work with data, to place current RDA activities in the overall landscape, to look at what other communities are doing in this area and to foster testing and adoption of RDA outputs. The goal is to identify common components and define their characteristics and services that can be used across boundaries in such a way that they can be combined to solve a variety of data scenarios such as replicating data in federations, developing virtual research environments, and automating regular data management tasks.[12]

The RDA Metadata Standards Catalog Working Group produced a machine-actionable catalog (MSC) of metadata standards originally submitted by all RDA WGs. The catalog system has an end-user input form and an API for submission from other software. The work builds on the outputs of the Metadata Standards Directory Working Group, which is responsible for creating the Metadata Standards Directory (MSD). Compared to MSD, the MSC offers improvements to the data structure of the records, an improved user interface, and the addition of the API. The catalogue is currently in

[8] Coen G: Introduction to Semantic Artefacts. [presentation] Presentation at the FAIRsFAIR workshop on semantics 22 Oct 2019. Available from: https://doi.org/10.5281/zenodo.3549375
[9] Shapes Constraint Language [web page] SHACL. Available from: https://www.w3.org/TR/shacl/
[10] Research Data Alliance [web page] Available from: https://www.rd-alliance.org/
[11] RDA Working Groups & Interest Groups [web page] Available from: https://www.rd-alliance.org/groups
[12] RDA: Data Fabric IG (DFIG) [web page] Available from: https://www.rd-alliance.org/group/data-fabric-ig.html

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

maintenance mode, but there is still minor development ongoing and the content continues to accumulate.[13]

The RDA Data Discovery Paradigms Interest Group (IG) currently has three Task Forces: Metadata Enrichment and Data/Metadata Granularity. The third task force is in the process of setting up a working group: Using schema.org for Research Dataset Discovery. There are also several pending RDA working groups in different states of activity, that touch upon metadata related questions: Research Metadata Schemas, Data Description Interoperability, and the Data Type Registries Working Group.[14]

2.1.1. Semantic interoperability

Interoperability has many levels as presented in figure 4 below. Semantic interoperability is, according to the research information standard dictionary CASRAI, the ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems. Semantic interoperability is achieved when the information transferred has, in its communicated form, all of the meaning required for the receiving system to interpret it correctly, even when the algorithms used by the receiving system are unknown to the sending system. Syntactic interoperability is a prerequisite to semantic interoperability.[15] It ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties, in other words 'what is sent is what is understood'. In the European Interoperability Framework (EIF), interoperability covers both semantic and syntactic aspects (European Union Directorate-General for Informatics, 2017). An interoperability framework specifies a set of common elements such as vocabulary, concepts, principles, policies, guidelines, recommendations, standards, specifications and practices.[16]

[13] RDA: Metadata Standards Catalog WG [web page] Available from: https://rd-alliance.org/groups/metadata-standards-catalog-working-group.html The cataloge can be found at https://rdamsc.bath.ac.uk/
[14] Research Metadata Schemas WG [web page] Available from: https://rd-alliance.org/groups/research-metadata-schemas-wg Data Description Registry Interoperability (DDRI) WG [web page] Available from: https://www.rd-alliance.org/groups/data-description-registry-interoperability.html
Data Type Registries WG & #2 https://www.rd-alliance.org/groups/data-type-registries-wg.html
[15] Semantic interoperability. [web page] CASRAI. [cited on 15.11.2019] Available from: https://dictionary.casrai.org/Semantic_interoperability
[16] ISA[2] [web page] EU. [cited 3.10.2019] Available from: https://ec.europa.eu/isa2/isa2_en

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

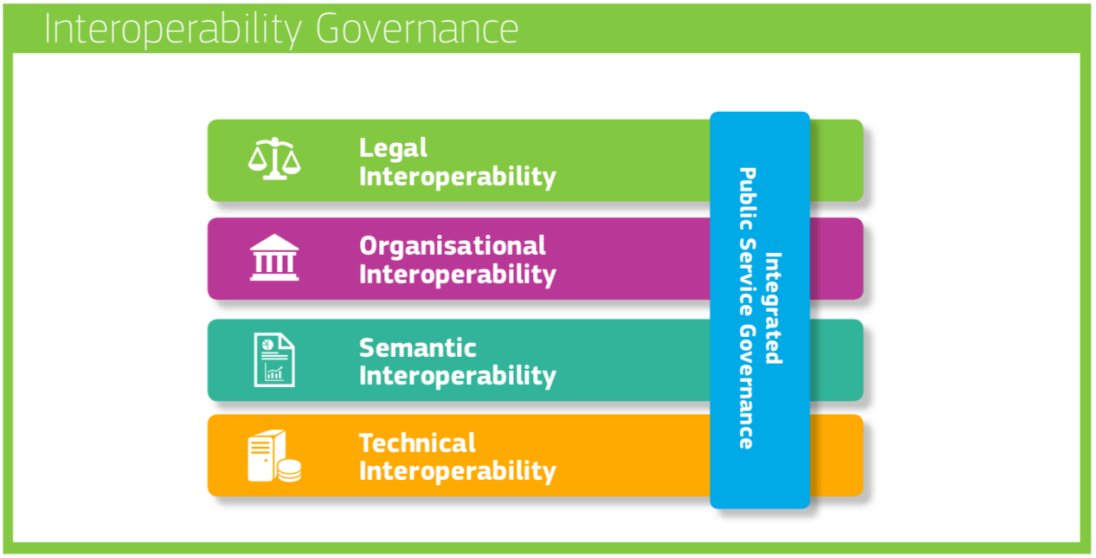

Figure 4. The European Interoperability Framework promotes seamless services and data flows for European public administrations

Semantic interoperability is, according to European Interoperability Reference Architecture (EIRA), about the meaning of data elements and the relationship between them. EIRA also defines Semantic Interoperability Agreements, which are an expression of consensus among a group of co-operation partners on the model and data entities that support common services. They include a developing vocabulary to describe data exchanges, and ensure that data elements are understood in the same way by communicating parties. The Semantic Interoperability Specification enables organisations to process information from external sources in a meaningful manner. Regarding research this process has to be done by the scientific communities, among researchers, but common principles and core elements can also be provided, which is one of the goals of the FAIRsFAIR project. The European work for interoperability sets a basis for research data interoperability, but is but far not enough for scientific use.

One of the challenging areas in semantic interoperability is trans-language interoperability, which requires multilingual semantic artefacts (eg. vocabularies, ontologies and concept schemes in different EU languages). This is a dimension that is especially important for humanities and social sciences, but should be considered as a generic point because it is important for open science and societal impact and outreach. Even though the English language has become a lingua franca within large parts of the STEM[17] domains, cultural and many vernacular contexts need to be actively included and integrated in the research discourse for research results and outputs to be disseminated and utilized outside research community. This is an issue where Europe can turn a challenge to a possibility and develop scientific resources that are available and reusable for people that don't speak English or want to integrate the data in contexts that are in other languages than English. Assessment of impact shouldn't be constrained to academic or even scientific impact. It can be a serious risk falling into the trap of thinking that all knowledge is in English. So, despite scientific

[17] STEM: Science, Technology, Engineering and Mathematics

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

context diversity, there is also cultural and linguistic diversity to manage. The EU terminology could also be developed by linking or extending it to the terminology of the research domain and EOSC.[18]

Semantic interoperability is, altogether, one of the important enabling elements of the FAIR principles (Guizzardi, 2019).

2.1.2. Semantic artefacts

Semantic artefacts are the tools which allow humans and machines to locate, access and understand (meta)data.[19] We use the term semantic artefact to bridge actual semantic problems regarding the use of terms like ontology, vocabulary and terminology within different communities. The term semantic artefact covers all of these (Figure 5).[20]

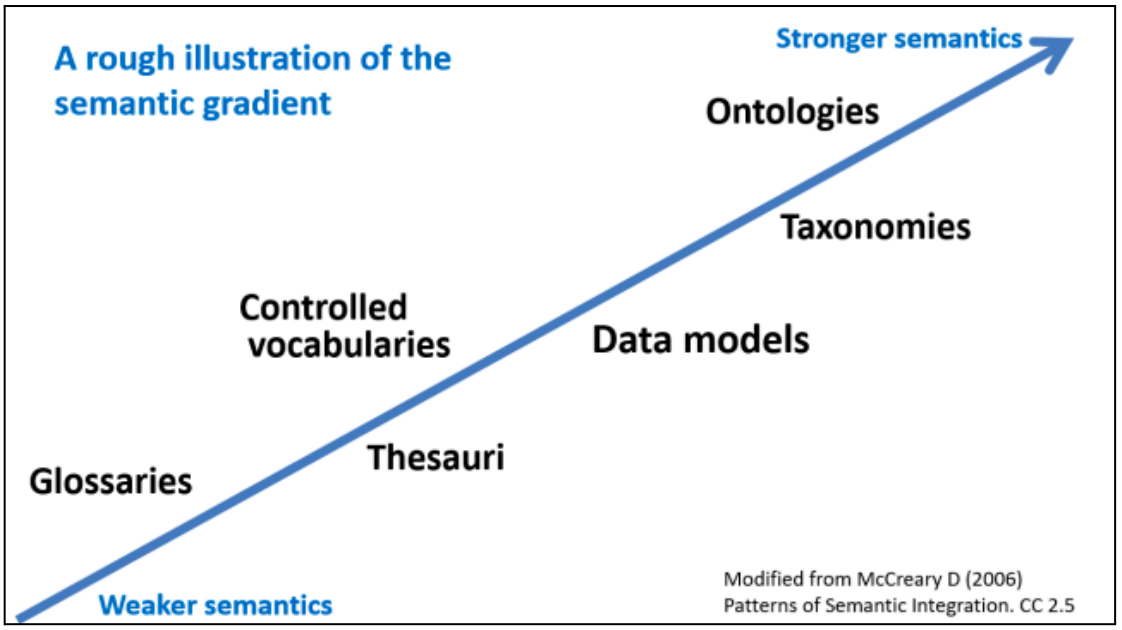

Figure 5: The term semantic artefact covers all steps on the ladder of semantic gradient.[21]

The availability and use of semantic artefacts has been pointed out as one of the key issues in a series of workshops on 'Services to support FAIR data'[22] and in the data collected for this report to

[18] IATE (Interactive Terminology for Europe). [web page] IATE. Available from: https://iate.europa.eu/home
[19] Gerard Coen: Introduction to Semantic Artefacts. [presentation] at Building the data landscape of the future, FAIRsFAIR workshop, Espoo, Finland, 22nd October 2019. https://doi.org/10.5281/zenodo.3549375. Adapted from: Leo Obrst "The Ontology Spectrum". Book section in of Roberto Poli, Michael Healy, Achilles Kameas "Theory and Applications of Ontology: Computer Applications". Springer Netherlands, 17 Sep 2010.
[20] This term has been suggested by the experts in our working group (T2.2) for a common terminology and has been used consistently in our work. However, the adoption of the term will need further negotiations, but these will take place within T2.2 and RDA VISSG and in FAIRsFAIR and hopefully the EOSC family. At this point we want to support this effort rather than resist it and therefore we use the term to achieve internal alignment within our project.
[21] Coen 2019. [presentation] https://doi.org/10.5281/zenodo.3549375.
[22] Services to Support FAIR Data: From Recommendations to Actions. OpenAIRE [Blog post] https://www.openaire.eu/blogs/2019-09-30-12-46-02 Final report is in progress.

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

promote the FAIR principles. According to survey data it was considered important to help service providers with interoperability. In practice this means integrating the artefacts into the workflows and tools like repositories. The data from FAIRsharing (Figure 6) shows how the Gene Ontology has achieved a strong position interlinking research data resources as a commonly used reference dataset.

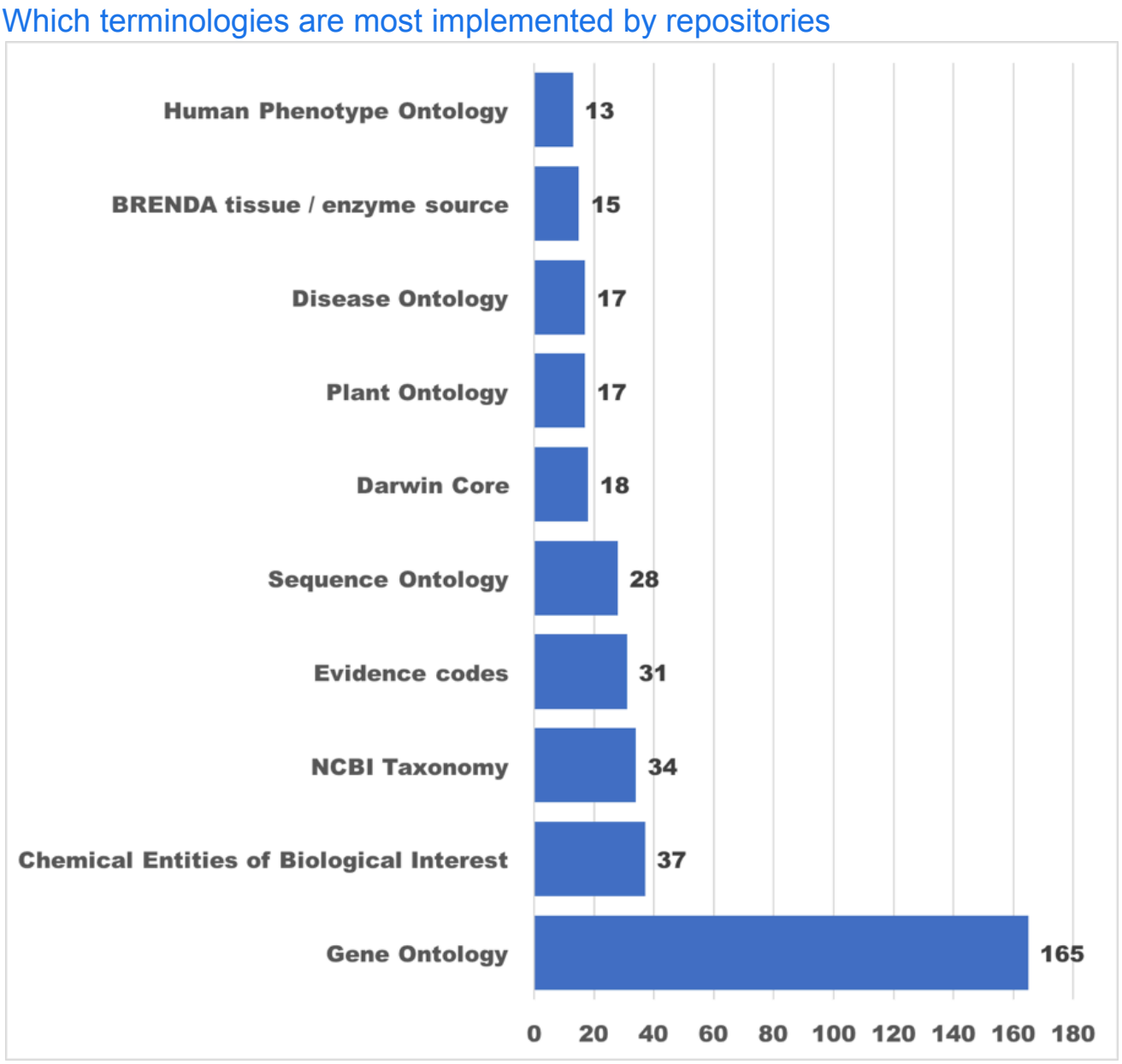

Figure 6. Many repositories already use terminologies, which is an efficient way to enhance FAIRness. Based on data from FAIRsharing. [23]

[23] McQuilton P. Bridging semantics and repositories. [presentation] at Building the data landscape of the future, FAIRsFAIR workshop, Espoo, Finland, 22nd October 2019. Data from FAIRsharing.

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

Based on the survey the following features are critical for the adoption of semantic artefacts:

1. Coverage in field (external):
    - They should be widely approved and adopted by the scientific community (indicator: use within community, mandates)

2. Coverage of content (internal):
    - They must cover a sufficient amount of the terminology needed (indicators: coverage, completeness and coherence).
    - They must have a structure that corresponds to the ontology of the field (indicators: certification, quality, community approval)

3. Governance (technical and legal):
    - They must be usable and fit the purpose (compatibility, format, granularity, workflow etc)
    - They must be actively maintained by a trusted, authoritative party (curation, versioning, persistence)
    - They must be open and documented

An important aspect is the findability of the semantic artefacts. There are both generic services for this like Bartoc registry, Linked Open Vocabularies (LOV)[24] and Industrial ontologies.[25] Several domains have mature services like the Ontology lookup service, BioPortal[26], Agroportal, BioPortal, and GEMET, but in other fields this is not an easy task and Google might be the only resort. The opinion that " … centralised registries of these semantic artefacts (trusted and FAIR) would be ideal" is not unusual. Also, beside the lack of semantic artefacts, there are some topics for which only proprietary ontologies are available. This can at least in theory restrict their use. The metadata of the semantic artefacts is important to enhance their findability. An ontology has been created to this end (MOT, Metadata for Ontology Description and Publication Ontology[27]). It has been implemented in the Agroportal and will be taken forward.

2.1.2.1. Interoperability for semantic artefacts

A workshop on semantics was arranged by FAIRsFAIR in October 2019[28] as a first effort within the project to discuss the FAIR principles in relation to interoperability. Many of the experts that participated work with linked data and well developed knowledge organisation systems and discussed the questions in the context of ontologies. The following thoughts were presented by some experts on interoperability

[24] Linked Open Vocabularies [web page] Computer Science School at Universidad Politécnica de Madrid. Available from: https://lov.linkeddata.es/dataset/lov
[25] Industrial Ontologies Foundry |web page] Available from: https://www.industrialontologies.org/
[26] BioPortal [web page] Available from: https://bioportal.bioontology.org/
[27] MOD, Metadata Vocabulary for describing and publishing ontologies. [git repository] https://github.com/sifrproject/MOD-Metadata-for-Ontology-Description-and-publication
[28] Building the data landscape of the future: FAIR Semantics and FAIR Repositories. Workshop on 22 October 2019 in Espoo, Finland arranged by FAIRsFAIR task 2.2 https://www.fairsfair.eu/events/building-data-landscape-future-fair-semantics-and-fair-repositories

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

- Should ontologies be aligned with upper-level ontology (e.g. BFO, DOLCE) as part of the FAIR maturity indicators? (I1)
- Implement ontology alignment (I1) (http://oaei.ontologymatching.org/2019/biodiv/index.html)
- Involve domain expertise in alignment validation (I1) (http://sws.ifi.uio.no/oaei/interactive/)
- Recommend the use of ontology design patterns (http://ontologydesignpatterns.org/wiki/Main_Page) and other shared best practices for ontology development (I1, I2)
- Identify the version of ontologies using unique permanent identifiers (PIDs/DOIs, https://w3id.org, PURL?) (I2)
- Use reification to overcome ontology mismatch (e.g. when searching across many datasets described using different conflicting ontologies) – attach attributes to the triples (I1, I3)

This work will continue within task 2.2. Furthermore, the RDA Vocabulary Services IG is planning new activities around semantic artefacts.[29]

The FAIRsharing database contains information about more than 700 terminology artefacts and the Bartoc service[30] counts them in thousands. To support a common and defined terminology, the RDA Data Foundations and Terminology IG continues creation of basic data concepts and framework models along with their vocabularies. The aim is to enhance synchronization of RDA conceptualization and enable better understanding within and between RDA groups.[31]

Multi-linguality is a both a challenge and an opportunity for the European digital research infrastructures. The illusion that all relevant knowledge is available in English (findable) or usable (impact) is not a good thing, but leads to dangerous monoculture as discussed above in the introduction about semantic interoperability.

The discussion on how semantic artefacts can be FAIR is ongoing in the FAIRsFAIR project and also in other contexts. As one interviewee said: “Those who say they are FAIR lie, it isn’t even defined properly yet”. The OBO principles[32] have been presented as a set of recommendations, that can support good practices. They are intended as normative for OBO Foundry ontologies.

[29] Vocabulary Services IG [web page] RDA. Available from: https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html. See also VSIG/VSSIG re-configuration [web page] RDA. Available from: https://www.rd-alliance.org/group/vocabulary-services-interest-group/post/vsigvssig-re-configuration [cited 22.11.2019]
[30] Basel Register of Thesauri, Ontologies & Classifications [web page] Basel University Library. Available from: https://bartoc.org/
[31] Data Foundations and Terminology IG. [web page] RDA. Available from: https://rd-alliance.org/groups/data-foundations-and-terminology-ig.html
[32] The OBO Foundry [web page]. OBO Foundry. [cited 22.11.2019] Available from: http://www.obofoundry.org/principles/fp-000-summary.html

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

2.1.2.2. CASE: OBO Foundry recommendation for ontologies

1. Open
The ontology **must** be openly available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed in altered form under the original name or with the same identifiers.

2. Common format
The ontology is made available in a common formal language in an accepted concrete syntax.

3. URI/Identifier space
Each class and relation (property) in the ontology must have a unique URI identifier. (The principle is to be reviewed)

4. Versioning
The ontology provider has documented procedures for versioning the ontology, and different versions of ontology are marked, stored, and officially released. (Exact wording also under review)

5. Scope
The scope of an ontology is the extent of the domain or subject matter it intends to cover. The ontology must have a clearly specified scope and content that adheres to that scope. (Work in progress)

6. Textual definitions
The ontology has textual definitions for the majority of its classes and for top level terms in particular. (To be reviewed)

7. Relations
Relations should be reused from the Relations Ontology (RO). (To be reviewed)

8. Documentation
The owners of the ontology should strive to provide as much documentation as possible. The documentation should detail the different processes specific to an ontology life cycle and target various audiences (users or developers). (Work in progress, more than 20 elements mentioned)

9. Documented plurality of users
The ontology developers should document that the ontology is used by multiple independent people or organizations.

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

10. Commitment to collaboration
OBO Foundry ontology development, in common with many other standards-oriented scientific activities, should be carried out in a collaborative fashion.

11. Locus of authority
There should be a person who is responsible for communications between the community and the ontology developers, for communicating with the Foundry on all Foundry-related matters, for mediating discussions involving maintenance in the light of scientific advance, and for ensuring that all user feedback is addressed.

12. Naming conventions
(Work in progress)

16.[!] Maintenance
The ontology needs to reflect changes in scientific consensus to remain accurate over time. (Work in progress)

These OBO Foundry ontology principles were formulated in the late 1990s to guide creation of new biomedical ontologies that followed the success of the Gene Ontology. They reflect the needs and limitations of an early, text-based ontology description format, OBO[33], that, most significantly, lacked means to enforce data typing. In practice, the integrity of datasets using OBO was ensured by the widespread use of publicly available editing tools that functioned as reference parses for this format. The alternative to reference parsers for text-based data formats it formally define them using context-free grammars that defines exhaustively all valid fields, their data types and cardinality. These grammars are usually represented in a Backus–Naur form (BNF) or one of its extensions[34], but they are seldom used outside computer science applications.

In the interviews, the need of good semantic artefacts was mentioned, also as a reflection on the State of Open Data Reports[35], as the most important single way to achieve good quality metadata and promote FAIRness. These tools can be integrated in the workflow in ways that make it possible to create interoperable (meta)data. In the survey, it was pointed out that "Two projects using DDI[36] might use different profiles and therefore only be partially interoperable". Standards are not enough, but the applications and implementation often requires human oversight and

[33] The OBO Flat File Format Specification [web page] Available from: https://owlcollab.github.io/oboformat/doc/GO.format.obo-1_2.html
[34] Extended Backus–Naur form. [web page] Wikipedia. [cited 22.11.2019] Available from: https://en.wikipedia.org/wiki/Extended_Backus%E2%80%93Naur_form
[35] The latest being The State of Open Data 2019. [report] Available from: https://www.digital-science.com/resources/portfolio-reports/the-state-of-open-data-2019/
[36] Data Documentation Initiative. [web pages] Available from: https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/overview

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

interpretation that also need to be much specification. An example of this is the standard for resource description and access designed by the library community.[37] On the other hand strict standardisation might also prevent innovation and the creation of new, rich data. Or as a representative from Figshare put it: "We need to be flexible and "vague" to serve all fields, but allow certain customising for organisations." This is of course the key use case that RDF/RDFS/OWL explicitly was created to address, by allowing the easy mash up of foundational generalized ontologies with more specific community and even organization-specific ontologies to create a resultant ontology which will achieve maximal interoperability corresponding to the intersection of terms of interest while being entirely unconstrained for narrower, more localized needs.

It is important to note, that technology will not in itself make data interoperable. Curation can help research and further discipline specific projects are needed to take things forward. The development should be driven by research. As mentioned in an interview, "using technology is not always the right way to solve a problem, we should ask: does this really help the researchers on their way? Only then is it a success."

Reference data is, according to ISA (Interoperability solutions for public administrations, businesses and citizens in the EU), a small, discrete set of values that are not updated as part of business transactions but are usually used to impose consistent classification. Reference data normally has a low update frequency. Reference data is relevant across more than one business system belonging to different organisations and sectors.[38] A reference dataset is a dataset that is used to collate, compare or normalise other data. Reference datasets play an important part in creating semantic interoperability and standardised metadata, in which case they can be called semantic artefacts. These offer controlled lists or identifiers to use in metadata..

By using shared semantic artefacts in metadata catalogs for datasets interoperability can be promoted. From a traditional dataset catalog perspective, an ontology can also be regarded as either a dataset itself (with metadata) or it can be used as a reference dataset if it provides identifiers and is properly defined. In the latter case they are used as lists or ontologies of accepted values in defined application profiles or core resources that offer persistent identifiers for linking data.

2.1.3. PIDs and PID services for research data

Content drift and link rot as menaces also for academic research publication and so called Cool URI's are not usually considered secure enough to ensure reproducibility of research. The Australian National Data Service (now part of Australian REsearch Data Commons) has produced some seminal guides on persistent identifiers.[39] According to their explanation, a persistent identifier is an

[37] Resource Description and Access (RDA). Available from: http://access.rdatoolkit.org/rdachp11-fi_rda11-1154.html
[38] ISA² Interoperability Training Course. 2014. Available from: https://joinup.ec.europa.eu/sites/default/files/document/2014-06/Semantic%20interoperability%20courses%20%20-%20Training%20Module%203%20-%20Reference%20data_v0.10.pdf, also see https://joinup.ec.europa.eu/svn/adms/ADMS_v1.00/ADMS_SKOS_v1.00.html
[39] Persistent identifiers: expert level. [web page] ANDS. [cited 13.10.2019] Available from: https://www.ands.org.au/guides/persistent-identifiers-expert

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

identifier which can be resolved to an appropriate representation of the resource (including downloading the resource itself, if it is online) (Figure 7). What makes them persistent is that they can be updated when (not if) the resource changes location or goes off-line, so it continues to resolve appropriately to a representation of the same resource. ANDS also introduced the term two-tiered systems to describe the method of achieving this machine actionable persistence over long term. By using services with dedicated web domain names as namespaces, like doi.org, the PID is resilient to changes in database technology or organisation structures or names. In practice, this requires a centrally managed redirect to a human readable web page that represents the content of the identifier and that offers a way to access the content, if it is digital. Under this arrangement, the URL may change as the object moves, but the identifier itself does not have to—so long as the URL resolution is kept up to date.[40]

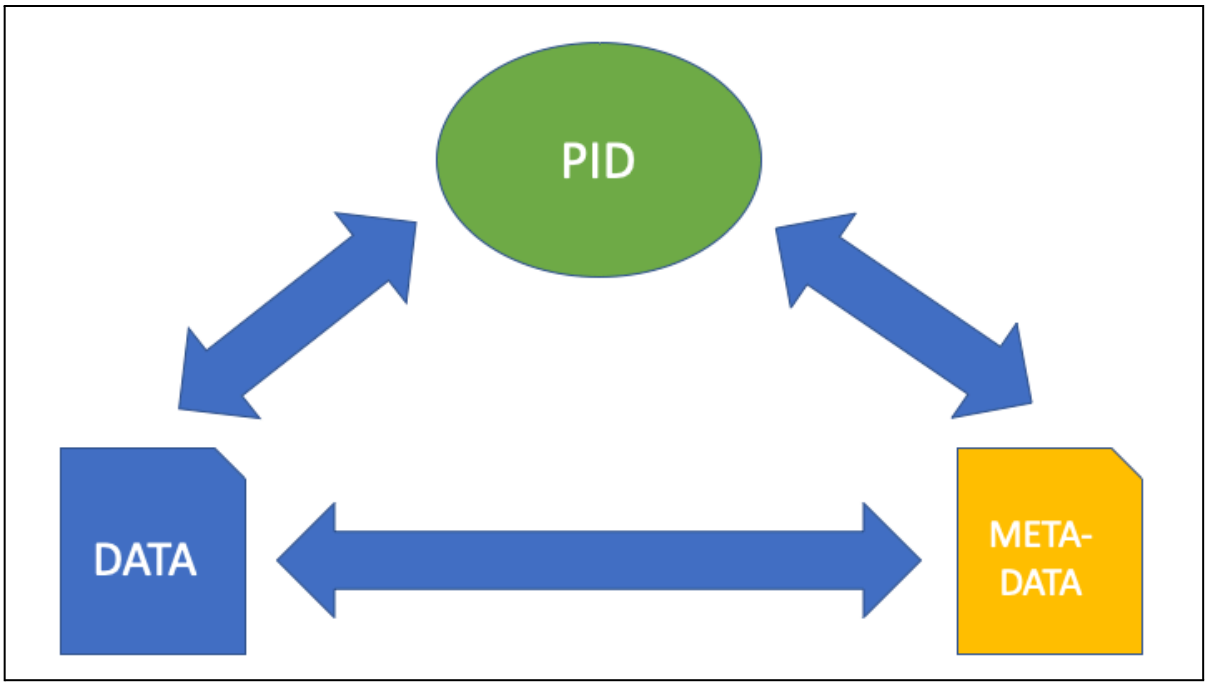

Figure 7: Relationships between PIDs, data and metadata. The resolver adds another layer to this model.

These two-tier systems have not inherently been accepted within the semantic web or LOD (Linked Open Data) communities, since it has been seen as an unnecessary layer when operating with machine actionable data. But there are different use cases and contexts also within the research community and sufficient nuance is necessary to meet different needs. Generally, for instance according to DataCite, good practice is to direct the human user to a landing page with metadata and licence information, when the represented object is a dataset. A persistent identifier meant for human users, for instance for data citation use, should be possible to identify as such. For example, Digital Object Identifiers (DOI) or Uniform Resource Names (URN) have distinctive syntax that makes it easy for a researcher to recognize and use in an appropriate way.

[40] Persistent identifiers: working level. [web page] ANDS. [cited 13.10.2019] Available from: https://www.ands.org.au/guides/persistent-identifiers-working

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

2.1.3.1. CASE: Recommendations about Persistent Identifiers

The RDA Data Fabric group has formulated recommendations about Persistent Identifiers as follows:

- A persistent identifier (PID) needs to be supported by a sustainable and trustworthy resolution system that will resolve PIDs to meaningful state information for machines and humans which are metadata attributes describing essential properties of a Digital Object (DO).
- A trustworthy PID resolution system needs to fulfil quality criteria still to be defined and needs to undergo regular quality assessment.
- The persistent PID record should be used to persistently bind the context of digital objects.
- A PID should be assigned to a Digital Object when it is registered at a trustworthy repository and thus becomes part of the domain of visible and findable data.
- A DOI should be registered when Digital Objects (data) are being published and citation metadata should be associated with it.[41]

Regarding formal dataset publication, the use of persistent identifiers is on a good way. But PIDs in search catalogues is only one use case, where DOIs are created for citability. It was pointed out in the interviews that there is also a need to create PIDs to support workflows and automation in metadata creation and machine actionable metadata at earlier stages of the data lifecycle. These PIDs will act as anchor points in the data lifecycle. The PID registries come into the picture at this stage, when a machine can act, maybe with the help of a generic protocol such as, e.g., the DOIP protocol, to enable intelligent data management services and repositories that can both create and act upon metadata with the help of PIDs. To support this work the RDA PID Kernel Information and Data Type Registries group work have provided RDA Recommendations (Weigel et al., 2015).

The DataCite DOI is an established solution for research dataset publication and can through close cooperation with other research information PID providers offer good value and has a strong brand that can support good data citation by uninitiated researchers. According to the interview, a data lifecycle perspective would be important and valuable also from a DataCite point of view. But DOIs for everything is not the answer in every situation. Variables and their PIDs might be of interest for DataCite, but PIDs for instruments might be more difficult to manage. How to get machine access to the data is an issue (but on the other hand DataCite only has aggregated metadata anyway).

[41] Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data, January 2017. [report] RDA Data Fabric IG. Available from: https://www.rd-alliance.org/group/data-fabric-ig/wiki/recommendations-implementing-virtual-layer-management-complete-life-cycle

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

PID saturation was not generally considered good. This is actually not in conflict with the point above, but the use of PIDs needs to be mindful and use cases should be clear and nuanced enough. The survey and desktop research confirmed that some fields have extensive use of PIDs. For instance, files, software, vocabularies, data models, concepts and other things might be referred to with some kind of persistent identifier. The most common PIDs for datasets seem to be the DOI, sometimes the URN, put also PURLs[42] and Handles[43] are in use in many contexts. Some communities have their own identifiers.

One use case for PIDs, not to be confused with the PID type registry above, that can support interoperability and machine actionability is the Data Type Registry (DTR), which can be used to register for instance:

A. how the various dimensions represented as variables in datasets of the form w1, d2, temp, etc., correspond to real world notions of weight, distance, temperature, etc.
B. what are the measurement units associated with each of those dimensions, e.g., Kelvin, Celsius, or Fahrenheit in the case of temperature.
C. how those dimensions are grouped or packed together in datasets.[44]

The challenge with both types of registries is that they should promote reuse rather than bulk creation of PIDs. To support interoperability they should be considered semantic artefacts and used mindfully.

2.1.3.2. CASE: DiSSCo

The Distributed System of Scientific Collections (DiSSCo RI) works for the digital unification of all European natural science assets and aims to make the data FAIR. The largest ever formal agreement between natural history museums, botanical gardens and collection-holding universities aims at transforming the fragmented landscape of natural science collections into an integrated knowledge base.

The possibilities that are created by linking for instance the 1 500 000 000 specimens with almost as many occurrence records and more than 40 000 datasets of the Global Biodiversity Information Facility GBiF opens important opportunities for research in the vicious challenges of our time.

In the Nordics, linked data is already an established way to manage taxon data and Swedish and Finnish taxons are

[42] PURL help [web page] PURL administration [Internet Archive]. Available from https://archive.org/services/purl/help
[43] HDL.NET® Information Services. [web page] Handle.Net Registry. Available from: https://www.handle.net/
[44] Data Type Registry. [web page] RDA. Available from: http://typeregistry.org/registrar/#

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

currently being linked via Taxonid[45] and accessible via their national portals. The Finnish Biodiversity Information Facility[46] offers its data in a well documented way, both for humans and machines and also offers a possibility to cite the dynamic dataset. The persistent identifiers are URI's according to CETAF recommendations.[47] The service also recently did a self evaluation on FAIR and the result was positive. The search for external mapping possibilities is ongoing although other data sources often provide scarce or no documentation nor machine actionable data.[48]

There still seems to be a certain push towards using two tier PID solutions and Handles will probably be introduced. Besides the Data Type Registry, Handles have been judged suitable. Recently, the following (all below) was presented by Alex Haridsty, expert from Cardiff University a propos DiSSCo:

- Options for Handles:

1. Acquire top-level prefix from an MPA – XX in XX.NNNNN/
2. Acquire second-level prefix – NNNNN in XX.NNNNN/

- From Crossref, Datacite, ePIC, etc. Ideally, 4 digits.

- Rejected options

1. Third level prefix e.g., from a Datacite member – too long!
2. International Geo Sample Number (IGSN) – assumes physical PID and digital PID are the same. Doesn't work for natural science specimens.

- Main considerations:
- Longevity/sustainability – 30 years at least
- Flexibility of metadata in PID (registry) records – need PID Kernel Information Profile for Digital Specimens

Further, thoughts on the future development covered other solutions that are discussed within the (GO) FAIR community:

- Digital Object Repositories
  - evolve from current repositories
- Digital Object Interface Protocol (DOIP)
  - specification exists, needs practical evaluation

[45] Taxonid [web page] Available from: http://taxonid.org/
[46] FinBIF. [web page]. Finnish Biodiversity Information Facility. Available from: https://laji.fi/en
[47] Güntsch A, Hagedorn G, Hyam R & Röpert D. CETAF stable identifiers for specimens. [poster] CETAF Available from: https://www.cetaf.org/sites/default/files/cetaf-istc_stable_identifiers_poster50x70.pdf
[48] Based on interview in September 2019.

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

- Digital Object Registries
  - overarching registries for searching
  - concept needs to be sharpened, relation with repositories
- Mapping/Brokering software and services
  - concepts, capabilities, implementations[49]

Recently, the DOI Foundation set up a filter and disabled the resolution of any PIDs not beginning with the '10.' string. This led to the immediate situation that all PIDs registered by other providers are not being resolved by using the doi.org resolver and are now forwarded to a landing page. Even though this was rolled back it shows that even two-tier systems are not always unproblematic, but require active management and have to be planned in a way that can support persistence over time.

Several of the interviewed experts mentioned that they see Artificial Intelligence helping in the future.

Identifiers.org provides resolution services to life science data and handles identifiers in the form of URIs and CURIEs. The PIDs consist of an assigned unique prefix and a local provider designated accession number (prefix:accession) and are thus an example of well-founded use of semantics in identifier syntax. The underlying Central Registry provides a centralized directory of these so called Compact Identifiers. Resource maintainers can use the Prefix Registration Request form to request a prefix in Identifiers.org for their databases or services. The California Digital Library's service Name to Thing (N2t) uses compact ID for several use cases.[50] This kind of logic is akin to that of Wikidata, building the identifier from an acquisition id to a URL. Also, the recently published draft for decentralized identifiers (DIDs) offers a type of identifier for verifiable, but decentralized digital identity. These new identifiers are designed to enable the controller of a DID to prove control over it and to be implemented independently of any centralized registry, identity provider, or certificate authority. The DID data model could in the future offer ways of creating or expressing identifiers in some use cases within research data management.[51] Ensuring semantic interoperability always will need active management and collaboration.

As a gap was identified between PID Information Types Recommendation and Data Type Registry Recommendation around what makes up PID Kernel Information, a new RDA WG was created to converge to the smallest number possible of versions (or profiles) of PID Kernel Information (one

[49] Hardisty A. DiSSCo Digital Specimens- Widening access to natural science collections. [presentation] Presentation at RDA GEDE Webworkshop Adaptation of Repositories to the Digital Object Interface Protocol on 22.5. 2019. Available at https://www.rd-alliance.org/group/gede-group-european-data-experts-rda/wiki/first-gede-do-workshop-september-18

[50] Documentation on Identifiers.org [web page] https://docs.identifiers.org/ and N2t [web page] https://n2t.net/e/compact_ids.html. [cited 21.11.2019]

[51] DID core. [web page] W3C. [cited 21.11.2019] Available from https://www.w3.org/TR/did-core/

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

was considered ideal but not likely).[52] The goal of the PID Kernel Information recommendation was to advance a small change to middleware infrastructure by injecting a tiny amount of carefully selected metadata into a Persistent ID (PID) record. This carefully chosen and placed information, targeted to internet scale services, is thought to have the potential to stimulate development of an entire ecosystem of third party services that can process billions of expected PIDs. This could be done with more information at hand about an object (no need for costly link following) than just a unique ID. The recommendation contains seven principles to enable machine actionable services. They state that the PID record should be a non-authoritative source for arbitrary metadata and stored directly at the resolving service.[53]

The purpose of the Persistent Identifier Interest Group in RDA is to synchronize identifier-related efforts, address important and emerging PID-related topics and coordinate activities, including appropriate RDA Working Groups, to practically solve PID-related issues from the engaged communities. It has almost 150 members.

The RDA Persistent Identification of Instruments working group (PIDINST) has collected use cases for persistent identification of instruments, and aims at aligning the collected metadata, and developing a metadata schema. In July 2019 the schema still contained a placeholder for the PID type as a suitable name for the instrument PID system still needs to be found.[54]

2.2. FAIR in the context of the Data Life cycle

In order to manage data throughout the research process, the documentation processes should be well established. There are several different models for the data lifecycle that define different stages of research. The community needs should be the guiding principle when creating solutions for data management and data citation. Raw data can be archived directly after its generation[55], but this is not always done. The generation of metadata and use of identifiers should be planned so that they support the workflows and need of the designated community. To do so, data needs to be extended with a minimal description which is useful for the scientist currently working on the data. The right kind of identifier should be allocated for different use cases (Figure 8).

[52] PID Kernel Information WG [web page] RDA. [cited 21.11.2019] Available from https://www.rd-alliance.org/groups/pid-kernel-information-wg
[53] Weigel T. et al. Recommendation on PID Kernel Information. [report] RDA; 2019. Available from: https://www.rd-alliance.org/group/pid-kernel-information-wg/outcomes/recommendation-pid-kernel-information
[54] Persistent Identification of Instruments WG. [web page] RDA. [cited 21.11.2019] Available from: https://rd-alliance.org/groups/persistent-identification-instruments-wg The Metadata schema is found at [git repository] https://github.com/rdawg-pidinst/schema
[55] Staiger C. FAIR data stewardship. [presentation] at Gov4Nano Kick-off meeting, March 2019. Available from: https://doi.org/10.5281/zenodo.2585691

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

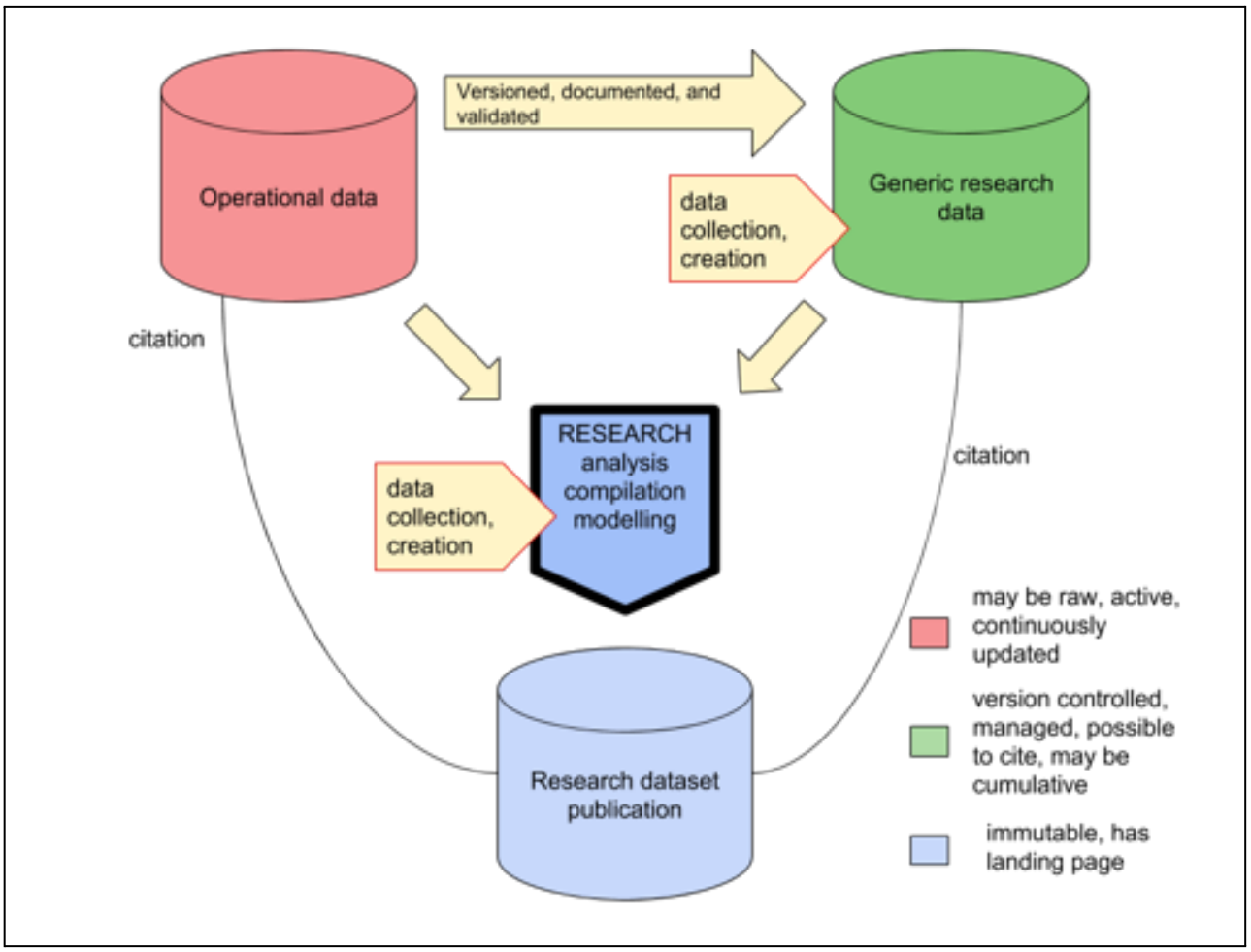

Figure 8: Supporting FAIR data: categorization of research data as a tool in data management (Parland-von Essen et al., 2018).

2.2.1. Data Repositories

A data repository is a service that is used to store and give access (with needed restrictions) to research data and metadata, is searchable and offers identifiers. A FAIR repository should serve both humans and machines. It is a solution that enables data services and data archives to store and share data. Data can be either datasets for research, semantic artefacts or code. A repository is not the same as a data set catalog (EIRA), which contains only metadata. Also indexing and search functionalities can be regarded as secondary in a strict interpretation of the concept of a repository and even metadata might technically be stored in separate services.

In domain independent repositories, research datasets are published as immutable datasets, with at least one data file, a landing page with generic metadata (for instance DataCite) and a persistent identifier for citation. They are often used by researchers that do not do research in fields with domain-specific repositories or formats for data or metadata. The service is most often used to publish datasets that are outcomes of answering a specific research question and whose main function is to underpin a research article or result and enable replication and citation for the researchers. The reuse value for secondary use is necessarily not great at the time of publication.

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

Metadata is usually created by the researcher and therefore often highly varying. These data repositories fulfil an important role in reporting to funders, offering metadata for aggregation, and linking data in services. The granularity of the dataset and metadata is generally on a low level and few services offer possibilities to add descriptive metadata to directories or files. Examples of this type of services are Figshare, Zenodo, EASY, Fairdata.fi. They might have more or less services like data curation or management on top of the repository. Some serve only as platforms for data owned by researchers or organisations with various data policies and practices, others might require a formal handover of rights and ownership to the provider.

Some digital research data infrastructures provide repositories that are aimed at certain research domains or communities. Examples include PANGAEA, Dryad or the ICOS carbon portal. In social sciences, services and data are offered by CESSDA members and in linguistics, by several language banks. These research data repositories have metadata formats and dedicated solutions that serve their designated community. Data is often published as immutable datasets.

The Research Data Repository Interoperability Working Group in RDA will establish standards for interoperability between different research data repository platforms. These standards may include (but are not limited to) a generic API and import/export formats.[56]

2.2.2. Evolving datasets and data citation

Research data is sometimes published and managed in databases, where data is published as individual nano publications and search queries might produce compiled datasets, which in turn can be given identifiers. Also queries can be stored and given persistent identifiers. This enables good prerequisites for replication and citation.[57] In practice dynamic and evolving dataset creates challenges to implementing the FAIR principles on data. DataCite gives four alternative ways to cite dynamic datasets, which offer different levels of reproducibility:

1. Cite a specific slice or subset
    - the set of updates to the dataset made during a particular period of time or to a particular area of the dataset
2. Cite a specific snapshot
    - a copy of the entire dataset made at a specific time
3. Cite the continuously updated dataset, but add Access Date and Time to the citation
    - Does not necessarily ensure reproducibility

[56] Research Data Repository Interoperability WG. [web page] RDA. [cited 21.11.2019] Available from: https://rd-alliance.org/groups/research-data-repository-interoperability-wg.html

[57] Cambridge Dictionary, vide “Repository” [web page] CUP. [cited 10.4.2019] Available from: https://dictionary.cambridge.org/dictionary/english/repository

Webopedia, vide “Repository”. |web page] Webopedia. [cited 10.4.2019] Available from: https://www.webopedia.com/TERM/R/repository.html

TeD-T, the Term Definition Tool, vide “Repository” |web page] RDA Data Foundation and Terminology Interest Group [cited 10.4.2019] Available from: https://smw-rda.esc.rzg.mpg.de/index.php?title=Repository

Data citation of evolving data. Recommendations of the Working Group on Data Citation. RDA;2015. Available from: https://rd-alliance.org/system/files/documents/RDA-DC-Recommendations_151020.pdf

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. Cite a query, time-stamped for re-execution against a versioned database[58]

The RDA Data Citation working group[59] produced a recommendation on evolving data in 2015. The solution comprises of the following core recommendations (Rauber et al., 2015):

- Data Versioning: For retrieving earlier states of datasets the data needs to be versioned. Markers shall indicate inserts, updates and deletes of data in the database.
- Data Timestamping: Ensure that operations on data are timestamped, i.e. any additions, deletions are marked with a timestamp.
- Data Identification: The data used shall be identified via a PID pointing to a time-stamped query, resolving to a landing page.

2.2.2.1. CASE: Evolving dataset citation[60]

Citing dynamic datasets can be done in different ways. The most thorough way is by creating a versioned database and storing the queries combined with use of persistent identifiers. This might be needed in some cases but open documentation of such solutions is not easy to find. Within the RDA work the adopters for evolving dataset citations are (the numbers indicate which RDA plenary they have been presented at):

Standards / Reference Guidelines / Specifications:
- Joint Declaration of Data Citation Principles: Principle 7: Specificity and Verifiability (https://www.force11.org/datacitation)
- ESIP Data Citation Guidelines for Earth Science Data Vers. 2 (P14)
- ISO 690, Information and documentation-Guidelines for bibliographic references and citations to information resources (P13)
- EC ICT TS5 Technical Specification (pending) (P12)
- DataCite Considerations (P8)

Reference Implementations
- MySQL/Postgres (P5, P6)
- CSV files: MySQL, Git (P5, P6, P8, Webinar)
- XML (P5)
- CKAN Data Repository (P13)

Pilot implementations, Use cases
- DEXHELPP: Social Security Records (P6)
- NERC: ARGO Global Array (P6)
- LNEC: River dam monitoring (P5)

[58] DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1. DataCite e.V., p .12. Available from: http://doi.org/10.5438/0014
[59] Data Citation WG [web page]. RDA. Available from: https://www.rd-alliance.org/groups/data-citation-wg.html
[60] This information was kindly compiled and provided by Andreas Rauber.

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

- CLARIN: Linguistic resources, XML (P5)
- MSD: Million Song Database (P5)

Adoptions deployed
- CBMI: Center for Biomedical Informatics, WUSTL (P8, Webinar)
- VMC: Vermont Monitoring Cooperative (P8, Webinar)
- CCCA: Climate Change Center Austria (P10/P11/P12, Webinar)
- EODC: Earth Observation Data Center (P14, Webinar)
- VAMDC: Virtual Atomic and Molecular Data Center (P8/P10/P12, Webinar)

In progress
- NICT Smart Data Platform (P10/P14)
- DendroSystem (P13)
- Ocean Networks Canada (P12)
- Deep Carbon Observatory (P12)[61]

Another approach is nanopublication for citing parts of datasets, sometimes referred to as micro attribution (Fabris et al., 2019). This has been applied in life sciences. According to nanopub.org[62] a nanopublication is a graph with three basic elements:

1. The Assertion: An assertion is a minimal unit of thought, expressing a relationship between two concepts (called the Subject and the Object) using a third concept (called the Predicate).
2. The Provenance: This is metadata providing some context about the assertion. Provenance means, 'how this came to be' and includes the methods that were used to generate the assertion and attribution metadata such as authors, institutions, time-stamps, grants, links to DOIs, URLs about the assertion.
3. The Publication Information: This is metadata about the nanopublication as a whole, and pertains to both the assertion and provenance. Similar to the provenance graph, the Publication Information contains "citation-like" metadata but pertains to the nanopublication and not just the assertion.

Documenting the research process and data provenance create needs for identifying workflows. The Common Workflow Language (CWL) (Amstutz et al., 2016) would also include manual activities.[63] There are different nascient ways of describing and structuring information about the processes and outputs of research relevant among these are the Research Object Crate[64] and on a higher level the RAiD[65].

[61] RDA Data citation WG. The webinars with all recordings, slides and links to papers are available at https://www.rd-alliance.org/group/data-citation-wg/webconference/webconference-data-citation-wg.html. [presentation] For presentation slides see https://www.rd-alliance.org/node/141/repository
[62] Nanopublication. [web page] Nanopub.org [cited 3.10.2019] Available from http://nanopub.org/wordpress/
[63] Common Workflow Language [web pages] https://www.commonwl.org/, https://w3id.org/cwl/v1.0 and [git repository] available at https://github.com/common-workflow-language/common-workflow-language
[64] Research Object Crate (RO-Crate) [web page] Available from: https://researchobject.github.io/ro-crate/
[65] Research Activity Identifier. [web page] RAiD. Available from: https://www.raid.org.au

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. The current status of FAIR data at a glance

3.1. International efforts to promote the FAIR principles

There are several stakeholders that strongly promote and advocate implementation of the FAIR data principles. Policy aspects will be covered in other deliverables of this project. Here, we will focus on the more technical aspects of this development. Central stakeholders are organisations like OECD[66], CODATA[67], WDS[68] and the Research Data Alliance[69], have long urged for data interoperability in science and similar technologies are also promoted outside the realm of research by organisations like Wikidata[70] and Open Knowledge[71]. The European Commission has promoted semantic interoperability at different levels, eg. through particular COST actions[72], Joinup and the ISA projects in the realm of PSI (Public Sector Information) and governmental data portals, and is also committed to the FAIR principles, being FAIR data one of the 8 challenges in the European Agenda for Open Science. This has been manifested in many funding decisions for research infrastructures and for instance in the seminal report and action plan "Turning FAIR into reality" (European Commission Expert Group on FAIR Data, 2018), a report from an EC High Level Expert Group, that is also the basis of this work. The executive summary gives four recommendations concerning the technical solutions presented as a "Case" below.

3.1.1. CASE: FAIR according to Turning FAIR into Reality

1. Central to the realisation of FAIR are **FAIR Digital Objects**, which may represent data, software or other research resources. These digital objects must be accompanied by persistent identifiers, metadata and contextual documentation to enable discovery, citation and reuse. Data should also be accompanied by the code used to process and analyse the data.

[66] Principles and Guidelines for Access to Research Data from Public Funding. OECD; 2007. [report] Available from: https://www.oecd.org/sti/inno/38500813.pdf
[67] CODATA mission. [web page] CODATA. [cited 3.10.2019] Available from: http://www.codata.org/about-codata/our-mission
[68] ICSU-WDS strategy 2019-2023. [report] Available from: https://www.icsu-wds.org/files/WDS_Strategic_Plan_2019-2023.pdf
[69] RDA [web page] Available from: https://rd-alliance.org/
[70] Wikidata. Data Access. [web page] Wikidata. [cited 3.10.2019] Available from: https://www.wikidata.org/wiki/Wikidata:Data_access#How_can_I_get_data_out_of_Wikidata?
[71] Vision and values. [web page] Open Knowledge Foundation. [cited 3.10.2019] Available from: https://okfn.org/about/vision-and-values/, Case studies [web page] Open Knowledge Foundation. [cited 3.10.2019] Available from: https://okfn.org/tools-services/case-studies/
[72] COST [web page] Available from: https://www.cost.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

2. FAIR Digital Objects can only exist in a **FAIR ecosystem**, comprising key data services that are needed to support FAIR. These include services that provide persistent identifiers, metadata specifications, stewardship and repositories, actionable policies and Data Management Plans. Registries are needed to catalogue the different services.

3. **Interoperability frameworks** that define community practices for data sharing, data formats, metadata standards, tools and infrastructure play a fundamental role. These recognise the objectives and cultures of different research communities. Such frameworks need to support FAIR across traditional discipline boundaries and in the context of high priority interdisciplinary research areas.

4. **FAIR must work for humans and for machines**: unlocking the potential of analysis and data integration at scale and across a distributed, federated infrastructure is one of the key benefits of making FAIR a reality.

Worth noting is also the larger European context of the twelve interoperability principles of the New EIF (European Interoperability Framework): Subsidiarity and proportionality, Openness, Transparency, Reusability, Technological neutrality and data portability, User-centricity, Inclusion and accessibility, Security and privacy, Multilingualism, Administrative simplification, Preservation of information and Assessment of Effectiveness and Efficiency which altogether fulfil the goals of Achieve Interoperability, and furthermore Achieve Legal Interoperability, Achieve Organisational Interoperability, Achieve Semantic Interoperability and Achieve Technical Interoperability.[73]

3.1.2. EOSC

The **European Open Science Cloud** (**EOSC**) is an European Commission initiative that started in 2015. It aims at developing a system of systems that can provide services to promote open science practices and enable access and reuse of research data. The European Open Science Cloud EOSC Portal was officially launched in November 2018. EOSC aims to support three objectives: (1) to increase the value of scientific data assets by making them easily available to a greater number of researchers, across disciplines (interdisciplinarity) and borders (EU added value) and (2) to reduce the costs of scientific data management, while (3) ensuring adequate protection of

[73] European Interoperability Reference Architecture (EIRA©) v3.0.0. P 65. Available from: https://joinup.ec.europa.eu/sites/default/files/distribution/access_url/2019-03/76cb237b-0de8-464c-84ca-1327945eac3e/EIRA_v3_0_0_Overview.pdf

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

information/personal data according to applicable EU rules. The FAIRsFAIR project is also an EOSC project.

EOSC currently supports the development of FAIR in various ways and through various approaches. There are many projects and working groups within the EOSC project ecosystem that work with landscaping activities and promote semantic interoperability. EOSC regional projects (EOSC Pillar, EOSC Synergy, ExPaNDS, EOSC Nordic and NI4OS-Europe) aim at connecting national initiatives, policies, infrastructure services and people to EOSC. The EOSC regional projects are a similar domain centric set of projects (ENVRI-FAIR, PaNOSC, ESCAPE, SSHOC and EOSC-Life). The ambition is to enhance the work on FAIR data practices and Open Science. On a common EOSC level the semantic interoperability work is only starting. The EOSC Executive Board has (currently) five working groups. The EOSC Governance includes the Governing Board, Executive Board and Stakeholder Forum. FAIRsFAIR has a Synchronisation Task Force working on providing FAIR-related support and input across all EOSC family related projects.

The EOSC Hub provides its users with a one-stop-shop for research data management due to a pooling effort of several service providers. The service providers are among others EUDAT CDI, the EGI Federation and INDIGO-DataCloud[74]. EUDAT CDI is also an infrastructure that enables allocating PID's, enables findability through B2FIND and even promotes semantic interoperability with the B2NOTE tool.[75] The EOSC Hub service catalogue includes 49 services. The EOSC Hub puts continuous effort into developing its services in close collaboration with entities such as GÉANT, OpenAIRE and RDA Europe. EOSC Hub is funded by the European Horizon 2020 research and innovation programme.

The pooling of efforts in the EOSC Hub services has made it possible to harvest from several sources via APIs, e.g. the EOSC Hub infrastructure itself is managed via Kubernetes APIs and the CloudFerro Data Collections Catalog is based on CKAN open source software and allows web APIs and the RESTful JSON API for access and discovery of datasets for several applications.[76]

3.1.3. FORCE11

FORCE11 is a self-organised community of scholars, librarians, archivists, publishers and research funders to facilitate the change toward improved knowledge creation and sharing. Members and sponsors of FORCE11 include commercial and non-profit publishers, libraries, scholarly societies, universities, other private and public sector organisations, and individual researchers, librarians, publishing professionals, corporate and public sector managers. The FAIR principles were born in this community, but it is more focused on research communication than on technical solutions or specifications. Recently, the maintenance of the FAIR principles for data was moved to GO FAIR.

[74] INDIGO-DataCloud [web page] Available from: https://www.indigo-datacloud.eu/
[75] B2NOTE. [web page] EUDAT. Available from: https://b2note.eudat.eu/
[76] EOSC Hub D7.2 First Report. 2018. [report] Available from. https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20D7.2%20v1%20Public.pdf

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

3.1.4. GO FAIR

GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the FAIR data principles. It offers an open and inclusive ecosystem for individuals, institutions and organisations working together through Implementation Networks (INs). GO FAIR promotes a minimal set of necessary protocols and standards and wants to support a wide variety of implementation choices for data, tools, and compute elements to participate in what they call the growing Internet of FAIR Data & Services (IFDS). The basic concepts are thus the Digital Object Model and the UPRI, a Unique, Persistent and Resolvable Identifier, that digital objects use. GO FAIR also stresses the need of very high quality, robust, and sustainable mapping services between UPRIs and human-readable terms that denote the same concept in digital objects. They call these semantic artefacts 'mapping tables' and point to them as critical infrastructure.[77]

With RDA GO FAIR has launched a "Metadata for Machines" workshop series (M4M) to assess the state of metadata practices in data-related communities and stimulate the creation and re-use of FAIR metadata standards and machine-ready metadata templates (definitions of metadata categories).

3.1.5. FAIRsharing

FAIRsharing is a community-driven resource and has a growing number of users, adopters, collaborators and activities[78], working to enable the FAIR principles. FAIRsharing is a large-scale service born from an early, community-driven portal launched in 2008 (MIBBI). FAIRsharing is hosted at the University of Oxford, and has close relations to CODATA, RDA, FORCE11 and other key stakeholders. Today the user base is a diverse set of stakeholders representing academia, industry, funding agencies, standards and research organizations, infrastructure providers and scholarly publishers—both national and domain-specific as well global and general organizations—involved in producing, managing, serving, curating, preserving, publishing or regulating data (Figure 9).

FAIRsharing also works as a service that provides content (metadata description) for a number of external tools, one example is the FAIR Evaluator tool (Wilkinson et al., 2019).[79] The joint RDA FAIRsharing WG resulted in an RDA Recommendation The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies.[80]

[77] The Internet of FAIR Data & Services. [web page] GO FAIR. [cited 3.10.2019] https://www.go-fair.org/resources/internet-fair-data-services/
[78] FAIRsharing communities. [web page] Available at: https://fairsharing.org/communities
[79] FAIR Evaluation Services. [web page] https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/
[80] FAIRsharing Registry: connecting data policies, standards & databases WG. [web page] Available from: https://www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html

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

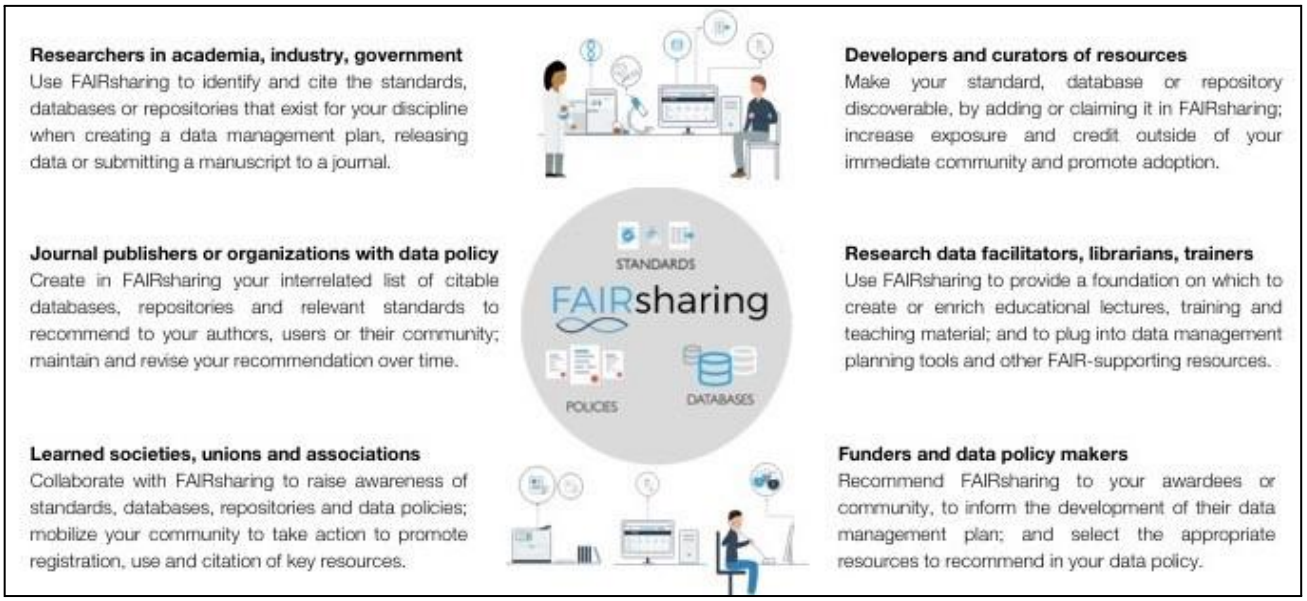

Figure 9: FAIRsharing aims at serving several key user groups.[81]

Using community participation, they curate information on standards employed for the identification, citation and reporting of data and metadata, via four standard subtypes: data policies, databases, standards and data collections. They also mint a DOI for each metadata record (Sansone et al., 2019). In addition, they recommend journals and publishers to encourage authors to cite the standards, databases and repositories they use or develop via the 'how to cite this record' statement, found on each FAIRsharing record, which includes a DOI. The recommendation also includes a notion that funders should recognize standards, databases and repositories as digital objects in their own right[81]

3.1.6. DataCite

DataCite is a leading global non-profit organisation in providing persistent identifiers (DOIs) for research datasets. Organizations within the research community join DataCite as members to be able to assign DOIs to their research outputs. This way, their outputs become discoverable and associated metadata is made available to the community through DataCite search and resolver services. DataCite develops additional services to improve the DOI management and findability, making it easier for their members to connect and share their DOIs with the broader research ecosystem and to assess the use of their DOIs within that ecosystem. DataCite is active in creating research information graphs and cooperates with ORCID, FREYA and other stakeholders.

3.1.7. re3data.org

[81] The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies. [report] RDA; 2019. Available from: http://dx.doi.org/10.15497/RDA00030

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

DataCite also maintains a Registry of Research Data Repositories, re3data.org.[82] This registry is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft). They describe data repositories and their data policies using an own schema if they fill these main criteria:

1. Run sustainably by a legal entity
2. Access conditions and terms of use must be clearly stated
3. Graphical user interface exists in English
4. Focus on research data

3.1.8. FREYA and Open Citations

FREYA is one of the projects funded by the European Commission under the Horizon 2020 programme. It was preceded by projects called THOR and ODIN. FREYA aims to extend the infrastructure for persistent identifiers (PIDs). FREYA wants to improve discovery, navigation, retrieval, and access to research resources. New provenance services are meant to enable researchers to better evaluate data and make the scientific record more complete, reliable, and traceable. By connecting new and existing PID services to make the most of the information available in different PID systems and promote the creation of a large graph of research information.[83] Research data and reference datasets can support this endeavour, but in itself FREYA is a potential tool for dissemination and further linking of the outcomes of a well functioning landscape of FAIR research data. The data model and work is closely related to work done within the RDA and also continues the extending use of JSON RESTful APIs[84] as a common technology for sharing and linking distributed information resources.[85]

This data is interesting for commercial actors, but openness is promoted by not only Horizon2020 programme financing, but also open data advocates like Open Citation[86] that also has introduced an Open Citation Identifier (OCI), which has a simple structure: the lower-case letters "oci" followed by a colon, followed by two numbers separated by a dash (e.g. oci:0301-03018). OCIs can be resolved using the OpenCitations OCI Resolution Service.

3.1.9. Research Data Alliance

Research Data Alliance (RDA) is a bottom-up, community driven global organisation that produces different types of research data management related solutions through its many working groups, interest groups, and other groups and networks. RDA has over nine thousand individual members and 58 organisational and affiliate members.

[82] re3data.org [web page] Available from: re3data.org
[83] Open science graphs are also developed in an RDA IG. Open Science Graph IG. [web page] Available from: https://www.rd-alliance.org/open-science-graphs-fair-data-ig
[84] JSON RESTful APIs [web page] RESTful API Tutorial. Available from: https://restfulapi.net/
[85] Introducing the PID Graph. [web page] FREYA.[cited 3.10.2019] Available from: https://www.project-freya.eu/en/blogs/blogs/the-pid-graph
[86] Open Citations was originally funded by JISC. [web page] Open citations. [cited 3.10.2019] Available from: https://opencitations.net/index

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

3.1.9.1. Interest groups

RDA Interest Groups (IGs) are networks and platforms for communication and coordination among individuals, outside and within RDA, with shared interests. IGs convene during the bi-annual RDA plenaries to discuss topical issues. Sometimes they also produce surveys, reports, and spin-off working groups (see next sub-chapter). IGs are long-term initiatives within the RDA and remain in operation as long as they are active, subject to periodic evaluation of their activity and its relevance to RDA aims. In October 2019 there were 55 interest groups.[87]

FAIRsFAIR partners have analysed the RDA Interest and Working Groups to identify those whose priorities most closely match the projects[88]. The IG's recognised in this exercise include RDA/WDS Certification of Digital Repositories IG, Repository Platforms for Research Data IG, Open Science Graphs for FAIR Data IG, Vocabulary Services IG, From Observational Data to Information IG, Data Policy Standardisation and Implementation IG, and Education and Training on Handling Research Data IG.

Of the above mentioned groups, the authors find the RDA/WDS Data Description Registry Interoperability IG[89] as having particular relevance. The International Science Council (ISC; formerly ICSU) used to have working groups for building an open scientific data catalog and a knowledge network. In 2017, it was decided within the World Data System (WDS) that the work should partly continue within the Research Data Alliance and as a cooperation with OpenAIRE and as a Scholix node.[90] The RDA work has resulted in the recommendation "Interlinking Method and Specification of Cross-Platform Discovery" (Aryani, 2018). This was, among other things, a precursor for the work on the PID registry (see the chapter on PID's).

3.1.9.2. Working groups

RDA working groups (WGs) aim at accelerating data sharing and exchange in concrete ways for specific communities. All WGs need to develop a recommendation in roughly 12-18 months time, i.e. over three bi-annual RDA plenary meetings. According to RDA guidelines, WG's should strive for

- elimination of a roadblock for data sharing,
- community specific substantive applicability (vs. universal applicability), and
- potential for quick adoption among active researchers.

Working Groups develop case statements describing the recommendation that the group will produce. The case statements go through community review and RDA Technical Advisory Board

[87] RDA in a Nutshell October 2019 [presentation| Available from: https://www.rd-alliance.org/sites/default/files/attachment/RDA-in-a-nutshell-October-2019.pptx
[88] "FAIRsFAIR Top RDA Working and Interest Groups" [web page] FAIRsFAIR, 2019. [cited 22.11.2019] Available from: https://www.fairsfair.eu/articles-publications/fairsfair-top-rda-working-and-interest-groups
[89] RDA Data Description Registry Interoperability WG [web page] Available from: https://rd-alliance.org/groups/data-description-registry-interoperability.html
[90] ICSU Knowledge Network Working Group [web page] ICSU. Available from: http://www.icsu-wds.org/community/working-groups/past-working-groups/knowledge-network

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

review. After that the RDA Council makes the final decision on whether to recognise and endorse the group. In October 2019 there were 28 working groups.[91]

The most relevant RDA WG for FAIRsFAIR work is the RDA FAIR Data Maturity Model Working Group, which is presented as a case below. The other relevant WG recognised by FAIRsFAIR partners (see above on RDA IG's) is Harmonizing FAIR Descriptions of Observational Data Working Group.

3.1.9.3. Other RDA groups

In addition to WG's and IG's there are coordination groups, and national and regional groups. From the point of view of FAIRness the most relevant is the RDA GEDE, which isn't a working group, but an invitation only coordination group. In practice it is a network of experts representing European e-infrastructures and e-infra related projects.[92] GEDE started operating in 2016. The first topic it tackled was PID's, by producing and publishing the document "Persistent identifiers: Consolidated assertion" (Wittenburg et al., 2017). At the 11th RDA Plenary in Berlin the GEDE launched a survey to recognise new topics. Four themes were decided as a result: 1) digital objects, 2) citing data, 3) digital repositories, and 4) blockchain technology.[93]

3.1.9.3.1. CASE: RDA FAIR Data Maturity Model Working Group

The RDA FAIR Data Maturity Model Working Group[94] aims at developing a discipline and data type agnostic common set of core assessment criteria for FAIRness.[95] The group also intends to create a generic and expandable self-assessment model for measuring the FAIRness related maturity level of a dataset. Group chairs are Edit Herczog and Keith Russell.

The group started by recognising indicators representing different aspects of FAIR data: what are to be evaluated to determine FAIRness? The next and currently on-going step is to put weight on those indicators. During this process the indicators will be grouped into three categories: 1) mandatory, 2) recommended, and 3) optional.

According to the preliminary results presented at the 14th RDA plenary meeting in Espoo, Finland, the group has recognised 53 FAIR data indicators, 15 of which have

[91] RDA in a Nutshell October 2019 [presentation] Available from: https://www.rd-alliance.org/sites/default/files/attachment/RDA-in-a-nutshell-October-2019.pptx
[92] GEDE Repository Topic Group [web page] Available from: https://rd-alliance.org/group/gede-group-european-data-experts-rda/wiki/gede-repository-topic-group
[93] GEDE - Group of European Data Experts in RDA [web page] Available from: https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda
[94] RDA FAIR Data Maturity Model WG. [web page] Available from: https://www.rd-alliance.org/groups/fair-data-maturity-model-wg
[95] FAIR Data Maturity Model WG: Case Statement. [report] [cited 22.10.2019] Available from: https://rd-alliance.org/group/fair-data-maturity-model-wg/case-statement/fair-data-maturity-model-wg-case-statement

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

been deemed mandatory, 30 recommended, and 8 optional. In the table below, we present only the mandatory indicators (as they were in October 2019), the full list can be found in the groups materials, linked to the group page on the RDA website. The number before the indicator gives the reference to which principle that particular indicator is connected to, as well as, whether the indicator targets data or metadata.

Findable	***Accessible***	***Interoperable***	***Reusable***
F1-01M Metadata is identified by a persistent identifier	A1-01M Metadata includes information about access conditions	I1-01M Metadata uses knowledge representation expressed in standardised format	R1-01M Sufficient metadata is provided to allow reuse, following domain/discipline-specific metadata standard
F1-01D Data is identified by a persistent identifier	A1-01D Data can be accessed manually (i.e. with human intervention)	I1-02M Metadata uses machine-understandable knowledge representation	R1.1-01M Metadata includes information about the licence under which the data can be reused
F4-01M Metadata is offered/published/exposed in such a way that it can be harvested and indexed [Priority]	A1-02M Metadata identifier resolves to a metadata record		R1.1-03M Metadata includes licence information in the appropriate element of the metadata standard used
	A1-03D Data identifier resolves to a digital object		
	A1.1-01M Metadata is accessible through a free access protocol		
	A1.2-01M Metadata includes information relevant for access control		
	A2-01M Metadata is guaranteed to remain available after data is no longer available		

Table 3: RDA FAIR Data Maturity Model Working Group

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

3.1.10. Other groups and stakeholders

Some countries and organizations interested in developing and deploying components of the Digital Object Architecture (DOA), including in particular the identifier/resolution mechanism founded DONA Foundation in Geneva, Switzerland in 2014 with the Corporation for National Research Initiatives (CNRI).[96] The Digital Object Architecture is an extension of the Internet architecture that consists of two protocols, the Digital Object Interface Protocol (DOIP)[97] and the Identifier/Resolution Protocol (IRP). The system consists of three components: an identifier/resolution service, a repository and a registry. It is built on TCP/IP protocols and bypasses the web protocol, which makes in an alternative to the REST API and other commonly used web technologies. The protocols are also being processed as RFCs as part of the Handle System. A reference implementation of the IRP is used for running Genetic Home Reference[98].

There are also other technical specifications that support interoperability created on behalf of different expert communities, like the Oxford Common File Layout[99] on a low level, or Frictionless data[100], Bagit[101] or METS[102] managed by the Library of Congress, that all seem to be quite widespread and in common use, to describe especially the structures of datasets (including metadata) and enable data transfers between systems and services. Using these kinds of open protocols and formats is an important part of the "A" and the "I" for many existing services, but they do not alone ensure semantic interoperability, only enable it.

The Wikimedia Foundation is also a relevant stakeholder because of Wikidata which offers identifiers, structured data and a SPARQL Endpoint. Also this data underpins at least partly the Google Knowledge Graph.[103]

For repositories, the Confederation of Open Access Repositories (COAR) is relevant, since they formulate a Pubfair specification for repositories[104] within their Next Generation Repositories initiative that also wants to include datasets. They have listed relevant API technologies and this way also promote several key recommendations.

---

[96] About DONA. [web page] The DONA foundation. [cited 22.11.2019] Available from: https://www.dona.net/aboutus
[97] The DOIP specification. [web page] The DONA foundation. 2018. Available from: https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf
[98] The IRP specification. [web page] The DONA foundation. [cited 21.11.2019] Available from: https://www.dona.net/specsandsoftware
[99] OCFL Specifications. [web pages] Available from: https://ocfl.io/, https://ocfl.io/0.3/spec/
[100] Frictionless data is a project by the Open Knowledge Foundation. [web page] Available from: https://frictionlessdata.io/
[101] Bagit is created by the Internet Engineering Task Force. [web page] Available from: https://tools.ietf.org/id/draft-kunze-bagit-16.html
[102] Metadata Encoding and Transfer Standard by the Library of Congress. [web page] Available from: http://www.loc.gov/standards/mets/
[103] The Google Knowledge Graph. [web page] Wikidata. [cited 22.11.2019] Available from: https://www.wikidata.org/wiki/Q648625
[104] Ross-Hellauer T, Fecher B, Shearer K, & Rodrigues E. Pubfair – A Framework for Sustainable, Distributed, Open Science Publishing Services. White Paper, Version 1 – September 3, 2019. [web page] COAR. [cited 3.10.2019] Available from https://comments.coar-repositories.org/

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

3.1.10.1. CASE: List of technologies monitored by COAR

- Activity Streams 2.0
- COUNTER
- Creative Commons Licenses
- ETag
- HTTP Signatures
- IPFS
- International Image Interoperability Framework
- Linked Data Notifications
- ORCID
- OpenID Connect
- ResourceSync
- SUSHI
- SWORD
- Signposting
- Sitemaps
- Social Network Identities
- Web Annotation Model and Protocol
- WebID
- WebID-TLS
- WebSub
- Webmention[105]

Schema.org[106] is an effort initiated by major search engine companies to add semantic tags in many available languages to web resources. The underlying ontology is maintained through collaborative community effort. The Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. Communities can define extensions to the ontology and approved for incorporation into schema.org by a committee if they are shown to be useful, needed, and widely in use.

Google is currently developing a Dataset Search[107], which will enable finding datasets stored across the web by way of a simple keyword search. The tool surfaces information about datasets hosted in thousands of repositories across the web, making these datasets more findable. Google uses schema.org in indexing and says that the more dataset repositories use schema.org and similar

[105] About technologies. [web page] COAR. [cited 22.11.2019] Available from: http://ngr.coar-repositories.org/technology/
[106] Schema.org [web page] Available from: https://schema.org/
[107] Google Dataset Search (beta) [web page] Google. Available from: https://toolbox.google.com/datasetsearch

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

standards to describe their datasets, the better visibility Google will give them. Code added to web resources can be validated using various tools, e.g. Google Structured Data Testing Tool[108].

The main impact of schema.org is to improve Findability of resources. BioSchemas[109] is an effort to extend schema.org with new types and properties useful to life sciences. Existing resources can have basic schema.org code generated using Bioschemas Generator.[110] The RDA working group on metadata also is planning a task force on schema.org.[111]

3.2. The landscape of digital research infrastructures

The landscape of infrastructures is diverse. There are domain agnostic and domain independent services like Zenodo or DataCite, EU funded common infrastructures such as EUDAT or OpenAIRE and commercial services e.g. Figshare. We also have a large abundance of infrastructures that are domain specific. Some are more based on expertise and others are created around instruments or data management services. This report focuses on large domain infrastructures, based on the ESFRI Roadmap 2018[112]. Still, it is important to acknowledge the role of the shared infrastructure in creating interoperability and sustainable technical solutions. We do not only have the EOSC; but also important partners and building blocks like FAIRsharing, Bartoc and other methods of creating semantic interoperability that are not even limited to research like the EIF (European Union Directorate-General for Informatics, 2017), ISA[2],[113] the Finnish Interoperability Workbench[114] or services for opening and linking data and managing persistent identifiers. These should also be taken into account to prevent creating silos between the research community and other domains of the society. Curated registries like the EOSC Hub, FAIRsharing and re3data.org are important resources for enabling implementation of the FAIR data principles.

The EUDAT service B2NOTE is a semantic data annotation service which can be integrated within the User Interface of any data repositories and services (Tomáš Kulhánek and Yann Le Franc, 2019). B2NOTE is integrated with the B2SHARE data and the community service Dendro (Karimova et al., 2017). Based on the W3C Web annotation model, B2NOTE provides the capability to link datasets or elements of datasets together with existing concepts/terms coming from ontologies/vocabularies without changing the underlying model of the data repository. To provide access to these concepts/terms to the user, a semantic index has been built. As of now, more than 5 million

[108] Google Structured data testing tool. [web page] Google. Available from: https://search.google.com/structured-data/testing-tool
[109] BioSchemas [web page]. ELIXIR [cited 22.11.2019] Available from: https://bioschemas.org/
[110] BioSchemas generator [web page] Available from: http://www.macs.hw.ac.uk/SWeL/BioschemasGenerator/
[111] Research Metadata Schemas WG [web age] Available from: https://www.rd-alliance.org/research-metadata-schemas-wg Meeting notes from RDA 14 Plenary [both cited 3.10.2019] Available from: https://docs.google.com/document/d/1UshlHlUPmV2FsLIOez8wLYnqYHCyqfyS-gnGIXLvcIw/edit
[112] ESFRI Roadmap 2018. [web page] ESFRI. [cited 3.10.2019] Available from: https://www.esfri.eu/roadmap-2018
[113] ISA² - Interoperability solutions for public administrations, businesses and citizens. [web page] EU. Available from: https://ec.europa.eu/isa2/home_en
[114] Yhteentoimiva Suomi offers tools for terminolgy work, reference data and data vocabularies for the Finnish government and is also used by the research data services provided by the Ministry of Education. [web page] Available from: https://yhteentoimiva.suomi.fi/en/

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

concepts, coming from Bioportal are available to the user. The extension of this semantic index to other domain's semantic artefacts remains a challenge as in many domains ontologies/vocabularies are hardly discoverable and interoperable.[115] To support such extension, it is necessary to establish a set of recommendations to support the creation of FAIR semantic artefacts, which is a question that will be addressed in the FAIRsFAIR project in cooperation with the RDA Vocabulary group.

Another service for semantic artefacts is the Basel Register of Thesauri, Ontologies & Classifications (BARTOC) produced by the Basel University Library, Switzerland. Its main goal is to list as many Knowledge Organization Systems from different subject areas as possible, in different languages and publication format, and any form of accessibility.[116] It is not as heavily dominated by life sciences as FAIRsharing and B2NOTE. The domain specific services will be touched upon below.

3.2.1. Energy

The field of energy research is very interdisciplinary, both because of societal significance and methodological and substance related issues. Questions related to energy are among the grand societal challenges of our time. For example affordable and clean energy is one of the United Nations sustainable development goals, but thematic is present in many of the other goals as well.[117] The research is partly related to research in physics and many other sciences, but has been categorized as a separate research domain in the ESFRI roadmap.The Energy ESFRI projects are EU-SOLARIS European Solar Research Infrastructure, the IFMIF-DONES International Fusion Materials Irradiation Facility and its DEMO Oriented NEutron Source, the MYRRHA Multi-purpose hYbrid Research Reactor, and the WindScanner European WindScanner Facility. Among landmarks ECCSEL ERIC and Jule Horowotz Reactor (JHR) are mentioned as landmarks in the ESFRI Roadmap 2018.[118]

Generally speaking, there is a lack of information in the Energy ESFRI's web pages relating to data management or FAIR principles. However, the WindScanner project has a work package dedicated to data processing and access management. They seek to address issues such as the means for enabling open access and e-science; procedures for data processing, validation and storage; and ways to preserve data integrity.[119] This seems to suggest that some FAIR aspects might be addressed e.g. findability through open access and interoperability through data validation.

---

[115] Goldfarb D & Le Franc Y, Enhancing the Discoverability and Interoperability of Multi-disciplinary Semantic Repositories, 2017. https://www.researchgate.net/profile/Yann_Le_Franc/publication/320058587_Enhancing_the_Discoverability_and_Interoperability_of_Multi-disciplinary_Semantic_Repositories/links/59cb8d260f7e9bbfdc3b38b5/Enhancing-the-Discoverability-and-Interoperability-of-Multi-disciplinary-Semantic-Repositories.pdf

[116] Bartoc. About. [web page] Universität Basel. [cited 21.11.2019] Available from: https://bartoc.org/en/content/about

[117] UN sustainable development goals. [web page] Available from: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[118] ESFRI roadmap. 2018, p. 61. [report] Available from: https://ec.europa.eu/info/sites/info/files/research_and_innovation/esfri-roadmap-2018.pdf

[119] Access. [web page] WindScanner Project [cited 21.11.2019] Available from: http://www.windscanner.eu/work-packages/work-package-5

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

3.2.1.1. Metadata

In the survey ISO 19115, Dublin Core Metadata, SKOS ontology, DCAT, PROV, OGC standards (sensorML, O&M) where mentioned. In FAIRsharing the Energy Industry Profile (EIP)[120] of ISO 19115-1:2014 is recognised as a relevant standard. It is an industry metadata exchange specification, but it is an open, non-proprietary exchange standard for metadata used to document information resources, and in particular resources referenced to a geographic location, e.g., geospatial datasets and web services, physical resources with associated location, or mapping, interpretation, and modeling datasets.

3.2.1.2. Semantic interoperability and artefacts

NVS, CheBI, PROV, SKOS, DCAT, DBpedia[121], schema.org but as mentioned above the relevant research or infrastructures are not necessarily always recognised by the researchers as being "Energy" labeled by the ESFRI forum. The relevant survey response came from a researcher working with EU-SOLARIS among other infrastructures and the respondent identified most closely with the Environment sector. This respondent was well-versed in semantic interoperability.

3.2.1.3. Identifiers

No persistent identifiers were mentioned in the survey data that was limited to the energy field.

3.2.1.4. CASE: Energy research

Some energy researchers experience limitations with respect to data management and publishing. In some situations, a PhD student will keep the data locally, making it difficult to obtain or reuse the data later on e.g. when the student graduates and leaves the institution.

There is a growing trend towards open science in publicly funded projects. However, projects that are privately funded usually come with data sharing restrictions because companies want to protect their competitive advantage.

"We mainly do experimental research where we are creating data that pretty much stays with us. At the moment we have no obligation, e.g from funders, to share data openly and therefore we have not yet studied this option. The data is stored in the research group folder. In some projects, we work with companies and then

[120] EIP specification. [web page] Energistics. [cited 21.11.2019] Available from: https://www.energistics.org/eip-specification/
[121] DBpedia [web page] Available from: https://wiki.dbpedia.org/

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

> the confidentiality of the research prevents the publication of all raw data with strict NDAs. We can only publish selected final results commonly agreed with the company." (Assistant Professor from a higher education institution in Finland)

3.2.2. Environment

The research in domains relevant for understanding the environment are gaining in societal and political importance as questions of climate change and its effects on biodiversity and ecosystems are becoming more evident to the public. The amount of data is large and typically there is a deluge of very diverse legacy data, and both long term data series and old taxonomies and specimens pose challenges for the modern user who wants to find and integrate this information with current research. These might not even be digitised or if they are digital, the formats might not be interoperable. Interoperability and metadata quality has been pointed out as an important challenge several years ago, but is still very much on the agenda.[122]

At the same time, immense amounts of new data are created through measurements and modelling. These require management of software and code. The realm of data is partly shared with other sciences, like the Life Sciences or other sciences, as for instance gene sequencing or biochemistry produce relevant data for understanding and describing the environment.

There are several ESFRI infrastructures that are creating interoperable data in these fields, and there are, despite considerable diversity, many good examples that are worth noting. One of the largest is the ENVRI-FAIR project, that made a landscaping analysis in 2015.[123] A landscaping effort on the interoperability of agricultural data was done within RDA in 2017.[124]

[122] Seys J et al. "Marine Data Management: we can do more, but can we do better?" [web page] IODE; 2006. [cited 22.11.2019] Available from: https://www.iode.org/index.php?option=com_content&view=article&id=3&Itemid=33. Salomon E. "Glances of the Landscape: Many environmental research infrastructures struggle with showing impact" [web page] RISCAPE; 2019. [cited 22.11.2019] Available from: https://riscape.eu/2019/03/21/glances-of-the-landscape-environmental-research-infrastructures/

[123] ENVRI-FAIR [web page]. Available from: http://envri.eu/envri-fair/. Especially the wiki on semantic interoperability has been of interest for this study. ENVRI wiki [cited 3.10.2019] Was available at https://wiki.envri.eu/display/EC/IC_11+Semantic+Linking+Framework

[124] Aubain S et al. Landscaping the Use of Semantics to Enhance the Interoperability of Agricultural Data. RDA Agrisemantics Working Group; 2017. [report] Available from: https://www.rd-alliance.org/system/files/documents/Deliverable1%20-%20Landscaping.pdf

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

Legacy data is a valuable part of environmental research resources because questions about long term development and change are important. DiSSCo, COST actions[125] and some RDA groups[126] have worked with creating guidelines and good practices for digitisation, which is an area that works with taxonomy and therefore also semantic artefacts. There are also efforts for integration of legacy data in eLTER[127], ACTRIS[128] and AnaEE[129] that have many archives and repositories with valuable, but diverse datasets. There is often a balance or compromise that is felt between the local context, needs and ways of work and the pressure to produce uniform data products. Therefore, there are different levels of data, ranging from raw or real time data and then data with different inputs of processing, cleaning, curation or formats. These levels of course are not equivalent.

Creating vocabularies and ontologies is one important strain of work, another is the development of common protocols and processes, that support creation of FAIR data. Also many databases and repositories use or seem to consider using PID systems.[130]

Some data intensive domains are quite mature with well documented data formats and services. Examples of these are Madrigal (EISCAT), PANGAEA (EMSO ERIC), some ACTRIS resources, EPOS, ICOS Carbon portal, EURO-ARGO, DEIMS (LTER), SeaDataNet CDI.

3.2.2.1. Metadata

Based on desk research, common schemas and data models are INSPIRE[131], Dublin Core, Darwin Core, Ecological Metadata Language and NetCDF. Other well documented examples of datasets that include metadata are GUISDAP and CEDAR. Geospatial information is often highly relevant and in the survey the ISO 19115 standard was mentioned several times.

3.2.2.2. Semantic interoperability and artefacts

[125] See COST actions. MOBILISE especially is focussed on creating interoperability in digitisation, Soil and Temperate Forests are other examples of actions for data integration. [Cited web pages 3.10.2019] Available from: https://www.mobilise-action.eu/ https://www.cost.eu/actions/CA18237/#tabs|Name:overview, https://www.cost.eu/actions/CA18207/#tabs|Name:overview

[126] The RDA interest group for agricultural data IGAD has started several efforts in form of WGs that directly focus on semantic interoperability by creating tools. [web page] See further web pages of IGAD, https://rd-alliance.org/groups/agriculture-data-interest-group-igad.html, AgriSemantics WG https://www.rd-alliance.org/groups/agrisemantics-wg.html, Capacity Development for Agricultural Data. https://www.rd-alliance.org/groups/capacity-development-agriculture-data-wg and the Wheat and RICE WG presented below.

[127] European Long-Term Ecosystem and socio-ecological Research Infrastructure [web page] Available from: https://www.lter-europe.net/elter

[128] European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases. [web page] Available from: https://www.actris.eu/

[129] Infrastructure for Analysis and Experimentation on Ecosystems. [web page] Available from: https://www.anaee.com/

[130] D2.1 survey data http://doi.org/10.5281/zenodo.3518922

[131] INSPIRE data models. Data specifications. [web page] INSPIRE. [cited 21.11.2019] Available from: https://inspire.ec.europa.eu/Data-Models/Data-Specifications/2892

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

Semantic artefacts that seem to be in quite wide use are published in the General Multilingual Thesaurus (GEMET). The obvious common factor in the environmental sciences is often geographic information, and so the INSPIRE format and ICSU-WDS cooperation with IODE and GEOSS give some relatively well defined layers of semantic interoperability. In fact, PANGAEA has even built as a prototype of the WDS data portal.[132] The AgroPortal is an important service for ontologies and great efforts have been made to support semantic interoperability both internally and with external resources (Jonquet et al., 2018).

Use of semantic web technologies is planned in EPOS and the ENVRI reference model provides a good basis for creating data vocabularies. In the eLTER projects there is a plan for creating standardised shared variables.

The RDA has several relevant groups that are looking for or promoting solutions that support interoperability. InteroperAble Descriptions of Observable Property Terminology Working Group (I-ADOPT WG) is not yet an endorsed working group. It aims at creating a community-agreed framework for representing observable properties by bringing together groups that have been working on developing terminologies to accurately encode what was measured, observed, derived, or computed. The consensus building will be informed by reviewing current practices and by a set of use cases, which will be used to define the requirements and to test and refine the common framework iteratively for data collected and created across the environmental sciences.[133] More than 50 people have announced interest as members. The RDA Interest Group on Agricultural Data (IGAD)[134] has also spurred interoperability work for important staple grains (wheat and rice).[135]

The RDA Agrisemantics working group[136] published a landscape report in 2017 which discussed semantic solutions and called for further development of the "open, persistent vocabulary for agriculture data and services", Global Agricultural Concept Space (GACS). Linked open data was seen as a solution to the challenges in findability.[137] The result will be discussed more closely below, when looking at the domain specific situation. The group also produced a recommendation titled "39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition" to promote the use of semantics.[138]

---

[132] Data Portal. [web page] World Data System [cited 21.11.2019] Available from: https://www.icsu-wds.org/services/data-portal

[133] RDA: InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT WG) [web page] RDA. https://rd-alliance.org/groups/harmonizing-fair-descriptions-observational-data-wg

[134] Agricultural Data Interest Group (IGAD) [web page] RDA. [cited at 21.11.2019] Available from: https://rd-alliance.org/groups/agriculture-data-interest-group-igad.html

[135] Wheat data Interoperability WG [web page] RDA. Available from https://www.rd-alliance.org/groups/wheat-data-interoperability-wg.html Rice data interoperabilitity WG [web page] RDA. Available from https://rd-alliance.org/groups/rice-data-interoperability-wg.html. [both cited 21.11.2019].

[136] AgriSemantics Working Group. [web page] RDA. [cited 21.11.2019] Available from: https://rd-alliance.org/groups/agrisemantics-wg.html

[137] Aubin S, et al. Landscaping the Use of Semantics to Enhance the Interoperability of Agricultural Data. RDA. [report] Available from: https://www.rd-alliance.org/system/files/documents/Deliverable1%20-%20Landscaping.pdf

[138] 39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition. AgriSemantics Working Group RDA; 2019. [report] Available from: https://doi.org/10.15497/RDA00036

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

3.2.2.3. Identifiers

The most commonly used persistent identifiers are DOI and URN. Where linked data solutions are in use, also cool URIs[139] and PURLs are prevalent, even if there seems to be a certain tendency towards introducing Handle System (see above the DiSSCo use case). Also, GenBank IDs and Wikidata IDs were mentioned as reference IDs.

3.2.2.4. CASE: AgroPortal

The AgroPortal was created in relation to the RDA Vocabulary and Semantic Services Interest Group. The important starting point was to use and integrate existing resources and ontology libraries and repositories constituted important source material. A distinction was made between metadata properties that are intrinsic to the ontology (name, license, description) and other information, such as community feedback or relations to other ontologies, which is information that an ontology library captures or creates. In the project, ontology metadata practices were studied by analyzing metadata annotations of 805 ontologies, reviewing the 23 most relevant vocabularies at the time are available for descriptive metadata for ontologies (including Dublin Core, Ontology Metadata Vocabulary, VoID), and comparing different metadata implementation in multiple ontology libraries or repositories. But the work didn't stop there. A new metadata model was created for the AgroPortal vocabulary and ontology repository, a platform dedicated to agronomy based on the NCBO BioPortal technology. The portal now includes 346 properties from existing metadata vocabularies that could be used to describe different aspects of ontologies: intrinsic descriptions, people, date, relations, content, metrics, community, administration, and access. (Jonquet et al., 2018)

3.2.3. Health & Food

The European research infrastructures focusing on food and health have a long history of efforts to share their data and increase the interoperability of different datasets. Another important aspect in

[139] Berners-Lee T. Cool URIs do not change. [web page] W3C. 1998. [cited 21.11.2019] Available from: https://www.w3.org/Provider/Style/URI.html

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

life science is the sensitivity of the data, since it might contain personal information of patients or be crucial for patenting. This is limiting the publication of the data, from raw data to integrated datasets.

In 2012, the EU project BioMedBridges[140] was launched[141] which focused on the development of necessary data infrastructure, including shared standards and semantic web technologies for medical research data. After its conclusion in 2015, the follow up project CORBEL[142] continues those efforts with having a dedicated work package "WP6 - Data access, management and integration" (Figure 10). "The planned services will benefit a range of users from biologists to software developers: for example, for identifying e.g. samples, generating data mappings to ontologies."[143]

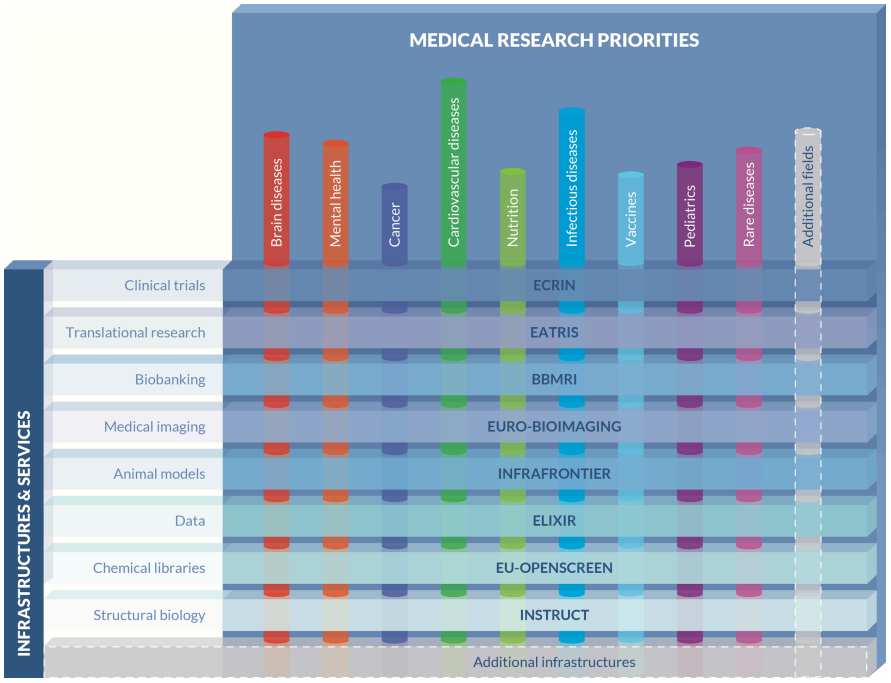

Figure 10: CORBEL WP3 aims to promote a transversal collaboration between RIs and medical research communities across borders and disciplines.[144]

Since March 1st 2019, 13 of the 16 ESFRIs are participating in the EOSC-Life cluster project of EOSC. The goal is to create an open collaborative space for digital biology[145] to be in line with the objectives of Open Science. In this project, the work package 6 "FAIRification and provenance

[140] Building data bridges from biology to medicine in Europe. [web page] BioMedBridges. [cited 21.11.2019] Available from: http://www.biomedbridges.eu/
[141] Final Report Summary - BIOMEDBRIDGES (Building data bridges between biological and medical infrastructures in Europe). Executive summary. [web page] CORDIS. [cited 21.11.2019] Available from: https://cordis.europa.eu/project/rcn/101852/reporting/en
[142] Coordinated Research Infrastructures Building Enduring Life-science services [web page]. CORDIS. [cited 21.11.2019] Available from: https://cordis.europa.eu/project/rcn/197885/en
[143] Data Access, Management and Integration. [web page] CORBEL. [cited 21.11.2019] Available from: https://www.corbel-project.eu/about-corbel/work-packages/wp6-data-access.html
[144] Medical/Translational Research Use Cases. [web page] CORBEL. [cited 21.11.2019] Available from: https://www.corbel-project.eu/about-corbel/work-packages/wp3-medicaltranslational-research-use-cases.html
[145] What is EOSC-Life? [web page] EOSC-Life. [cited 21.11.2019] Available from: http://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

services" has the goal to "enhance interlinked repository of registries and identifiers, as a common basis of metadata models and interoperability in the EOSC-integrated data sets". Here especially Task 6.2: "Identification and application of registries for FAIR data infrastructures" with the focus on identifying and interconnecting registries with common metadata models might be a good opportunity for collaboration with WP2 in FAIRsFAIR.

ESFRI	EOSC-Life	BioMedBridges (until 2014)	CORBEL
AnaEE			
BBMRI ERIC	X	X	X
EATRIS ERIC	X	X	X
ECRIN ERIC	X	X	X
ELIXIR	X	X	X
EMBRC ERIC	X	X	X
EMPHASIS	X		X
ERINHA	X	X	X
EU-IBISBA			
EU-OPENSCREEN ERIC	X	X	X
Euro-BioImaging			X
INFRAFRONTIER	X	X	X
INSTRUCT ERIC	x	X	X
ISBE	x		X
METROFOOD-RI			
MIRRI	X		X

Table 4: Membership of ESFRI in EOSC-Life, BioMedBridges and CORBEL

The W3C Semantic Web Health Care and Life Sciences Interest Group (HCLS IG)[146] delivered high level and architectural vocabulary for example the Translational Medicine Ontology (TMO) (Denney et al., 2009). The group was discontinued in 2018 and the work continued in Semantic Web Health Care and Life Sciences Community Group (HCLS CG)[147].

[146] Semantic Web Health Care and Life Sciences Interest Group. [web page] W3C. [cited 21.11.2019] Available from: https://www.w3.org/blog/hcls/
[147] Semantic Web in Health Care and Life Sciences Community Group. [web page] W3C. [cited 21.11.2019] Available from: https://www.w3.org/community/hclscg/

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

Even though various groups seem to work on standards, no common metadata standard could have been identified within the research field at large. This might be caused by the large diversity of research methods and observables and due to the complexity of the different sub fields. One example is the Minimum Information About Biobank data Sharing (MIABIS) Community Standard, which aims to standardize data elements used to describe biobanks[148], research on samples and associated data. Similar standards might exist in other subfields, but were not findable on the public pages of the RIs.

The collaboration of most research projects in minimum one RI and therefore also the inclusion in the EOSC-Life cluster might help to ensure a full interoperability also between the subfields.

ELIXIR[149] - the European life-science infrastructure for biological information - can be identified as a leading RI promoting Europe-wide standards that can be used to describe life science data. It has launched the Interoperability Platform[150] to help people and machines to discover, access, integrate and analyse biological data (Figure 11).

In the platform four tasks are established which work on FAIR Service Architecture (Task 1), Interoperability with a Purpose (Task 2), Capacity Building (Task 3), Interoperability Services for the Cloud (Task 4). Furthermore, ELIXIR has a Bioschemas group, which extended the Schema.org[151] specifications and definitions to the Life Sciences and aims to support the usage of Bioschemas.

[148] Minimum Information about Biobank Data Sharing. [git repository] https://github.com/MIABIS/miabis/wiki
[149] ELIXIR. [web page] ELIXIR. [cited 21.11.2019] Available from https://elixir-europe.org
[150] Interoperability Platform [web page] ELIXIR. [cited 21.11.2019] Available from: https://elixir-europe.org/platforms/interoperability
[151] Bioschemas. [web page] ELIXIR. [cited 21.11.2019] Available from: https://elixir-europe.org/platforms/interoperability/projects/bioschemas

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

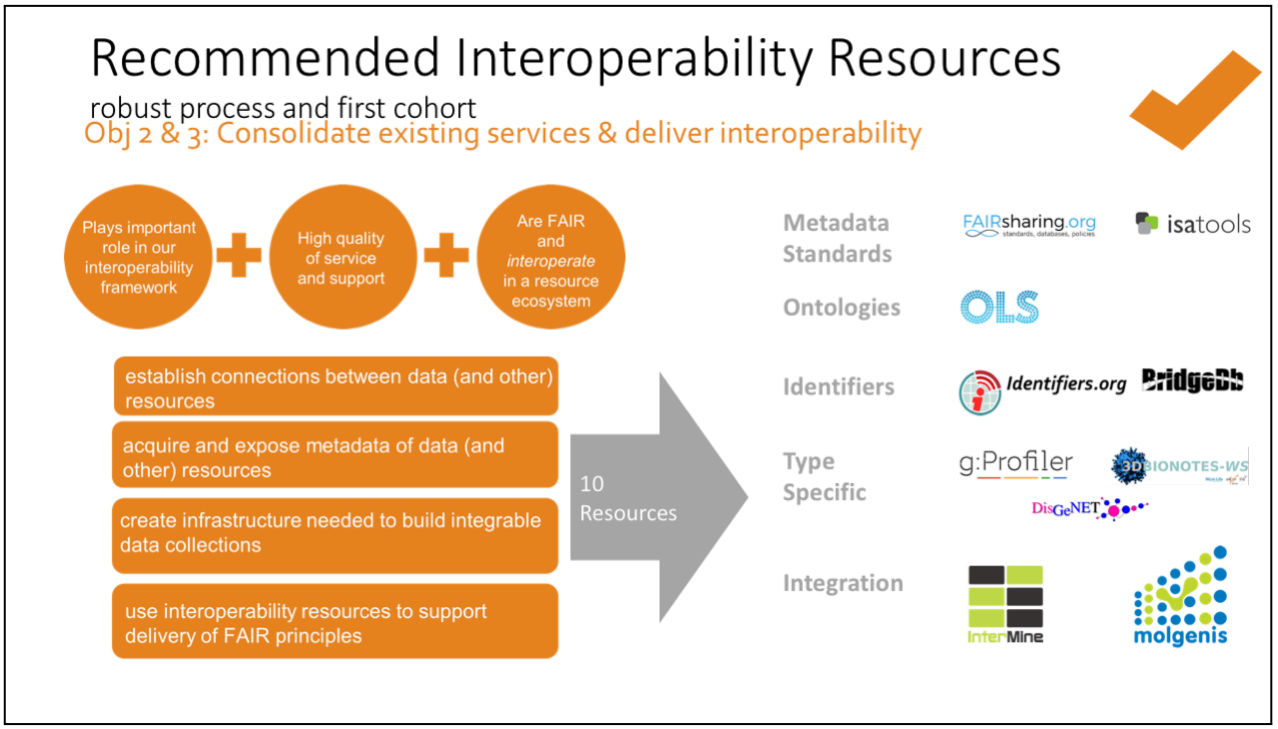

Figure 11: ELIXIR interoperability platform[152]

Overall the Life Sciences are working on the creation of complete interoperability between their RIs and have with EOSC-Life and CORBEL two projects which have interesting potential for collaboration with FAIRsFAIR. One big challenge is how to define FAIR between the biobanks and other lifesciences, i.e. how to integrate biomedical research and clinics.

3.2.3.1. Metadata

In the survey, mentioned metadata schemas were Dublin Core and DataCite schema. Many times the participants of the survey did not mention a metadata standard but a data standard or an ontology. In the collection of FAIRsharing[153], 785 standards are mentioned. These include both data standards and metadata standards.

3.2.3.2. Identifiers

The persistent identifiers mentioned in the survey include DOI, URN, Handle, PURL, PubMedID, PMC as well as CETAF identifiers. With respect to persistent identifiers, ELIXIR refers to a published review (McMurry et al., 2017). This review lists desirable characteristics for database identifiers in the life sciences. The overall conclusion of the paper agrees with the analysis of the available information for all Health & Food ESFRIs.

[152] Interoperability Platform. [web page] ELIXIR [cited 21.11.2019] Available from: https://elixir-europe.org/platforms/interoperability
[153] [Stats] [web page] FAIRsharing. [cited 21.11.2019] Available from: https://fairsharing.org/summary-statistics/?collection=all

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

3.2.4. Physical Sciences & Engineering

The ESFRI roadmap 2018 report identified three thematic sub-areas within the Physical Sciences and Engineering (PSE) Domain; and their corresponding RIs, as shown in the table below:

Astronomy and Astroparticle physics	**Particle and nuclear physics**	**Analytical physics**
• Landmark SKA (Square Kilometre Array). • Landmark ELT (Extremely Large Telescope) • Landmark CTA (Cherenkov Telescope Array) • Project KM3NeT 2.0 (KM3 Neutrino Telescope 2.0) • Project EST (European Solar Telescope)+	• Landmark HL-LHC (High-Luminosity Large Hadron Collider) • Landmark FAIR (Facility for Antiproton and Ion Research) • Landmark SPIRAL2 (Système de Production d'Ions Radioactifs en Ligne de 2e génération) • Landmark ELI	• Landmark ESRF EBS (European Synchrotron Radiation Facility Extremely Brilliant Source) • Landmark European XFEL (European X-Ray Free-Electron Laser) • Landmark ILL (Institut Max von Laue-Paul Langevin) • Landmark European Spallation Source ERIC • Landmark EMFL

Table 5: RIs within the PSE thematic groups

Generally, ESFRIs in the PSE domain seem to recognize the value of open science. To this end, a number of them (Project KM3NeT 2.0, Landmark CTA, Landmark EMFL, Landmark ESRF EBS, Landmark European XFEL, & Landmark ILL) have enacted data management plans or policies with the aim to incorporate FAIR principles. Joint projects have been formed in order to address common challenges across the PSE ESFRIs.

The Astronomy ESFRI & Research Infrastructure Cluster project (ASTERICS) developed the cross-cutting synergies and common challenges shared by astronomy, astrophysics and astroparticle ESFRIs: ELT, SKA, and CTA, and KM3NeT 2.0, with liaison building up with the ESFRI Project EST.[154] It was made up of five work packages, two of which related to data management and interoperability (OBELICS & DADI). OBELICS (OBservatory E-environments LInked by common ChallengeS) had the goal of enabling interoperability and software reuse for the data generation, integration and analysis of the ASTERICS ESFRI and pathfinder facilities. There was a focus on open standards, software libraries and unified solutions for data processing across huge sophisticated databases.

Data Access, Discovery and Interoperability (DADI) aimed at improving data availability, discovery and usage through an interoperable system that is easily accessible.[155] Part of the work from

[154] ASTERICS wiki pages. [web page] 2017. Available from: https://www.asterics2020.eu/dokuwiki/doku.php?id=open:gen:start
[155] WP4: Data Access, Discovery and Interoperability (DADI). [web page] ASTERICS. [cited 21.11.2019] Available from: https://www.asterics2020.eu/dokuwiki/doku.php?id=open:wp4:start

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

ASTERICS will continue within ESCAPE i.e. European Open Science of Astronomy & Particle Physics ESFRI research infrastructures (1 February 2019 - 31 August 2022) As noted in the final report, "the results and legacies of ASTERICS DADI will also be used by the WP4 of the ESCAPE Cluster, CEVO – Connecting ESFRI projects to the EOSC through the Virtual Observatory, which includes a task on FAIRisation of ESFRI data."[156]

H2020-ESCAPE aims to address the Open Science challenges shared by ESFRI facilities (CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA) as well as other pan-European research infrastructures (CERN, ESO, JIV-ERIC, EGO-Virgo) in astronomy and particle physics research domains.[157] The ESCAPE project aims to have deliverables that are strongly connected to EOSC in terms of management, governance, e-infrastructures, services etc. Additionally, it has a goal to build up a federated, sustainable infrastructure based on FAIR principles.[158]

The Photon and Neutron Open Science Cloud (PaNOSC) brings together six strategic European RIs (ESRF, CERIC-ERIC, ELI Delivery Consortium, the European Spallation Source, European XFEL and the Institut Laue-Langevin – ILL), and the e-infrastructures EGI and GÉANT. It's main goal is to contribute to the construction and development of EOSC - hence providing researchers with a single access point to universal and cross-disciplinary data. The project collaborates with other related European partners to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services.[159] The project has a data policy framework that aims at increasing awareness about Open Data and building best practices that support FAIR principles.[160]

	PaNOSC	ESCAPE
CTA	x	x
KM3NeT	x	x
ELT	x	x
EST	x	x
SKA	x	x
ELI	x	
ESRF EBS	x	

[156] ASTERICS. Astronomy ESFRI & Research Infrastructure Cluster. Part B, 3rd Periodic Report. [report] Available from https://www.asterics2020.eu/dokuwiki/lib/exe/fetch.php?media=open:wp1:2019_technical_report_part_b_v1.0.pdf
[157] ESCAPE Summary for Press Release. Tuesday 20 Nov 2018. [report] Available fom https://indico.in2p3.fr/event/18279
[158] About us. ESCAPE. [web page]. ESCAPE. [cited 21.11.2019] Available from: https://projectescape.eu/about-us
[159] About PaNOSC. PaNOSC. [web page] PaNOSC. [cited 21.11.2019] Available from: https://www.panosc.eu/about-panosc/
[160] PaNOSC data policy framework. [web page] PaNOSC. [cited 21.11.2019] Available from: https://www.panosc.eu/data-policy/panosc-data-policy-framework/

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

ESS ERIC	x	
XFEL	x	
ILL	x	
FAIR		x
HL-LHC		x
EMFL (European Magnetic Field Laboratory)		

Table 6: RIs in PaNOSC and ESCAPE

ExPaNDS is an EOSC Photon and Neutron Data Services project with the aim to expand, accelerate and support data management and data services provided through EOSC for major national Photon and Neutron Research Infrastructures (PaNRIs) in delivering world-leading science[161]. It brings together 10 European national infrastructures from the Photon and Neutron research domains[162] and aims at providing scientific users with EOSC services built upon FAIR principles. ExPaNDS will consolidate data and software services suitable for a wide range of users in the PaN ecosystem. The project plans to collaborate with other projects (e.g. PaNOSC, EOSC Synergy, EOSC Pillar, EOSC Nordic, NI4OS-Europe, etc) in strengthening and realizing open science through FAIRification of data.[163] One of its objectives is to adopt the FAIR data certification scheme, under development within FAIRsFAIR.[164]

With the exception of EMFL (European Magnetic Field Laboratory), all the other ESFRIs belong to either ESCAPE, PaNOSC or both. However, EMFL has a data management policy which aims at supporting FAIR data principles. ExPaNDS is a new project that has links to PaNOSC. Both PaNOSC and ESCAPE are fairly new and have data policies geared towards implementing FAIR principles. Therefore, it would be beneficial to collaborate with them. WP2 of FAIRsFAIR could provide guidance towards technical implementations of semantic interoperability. Due to the wide coverage of ESFRIs within PaNOSC and ESCAPE, it might be beneficial to work with a few select ESFRIs at the beginning.

3.2.4.1. Metadata

[161] ExPaNDS project website [web page] ExPANDS [cited 21.11.2019] Available from: https://expands.eu/about-expands/
[162] PANOSC project website [web page] [cited 21.11.2019] Available from: https://www.panosc.eu/related-projects/expands/
[163] ExPaNDS kick off 2019 Meeting [web page] [cited 21.11.2019] Available from: https://indico.desy.de/indico/event/23649/overview
[164] ExPaNDS Project overview [web page] ExPANDS [cited 21.11.2019] Available from: https://www.eosc-hub.eu/sites/default/files/ExPaNDS.pdf

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 are continuous efforts to include metadata descriptions for datasets and to better manage the metadata in a more FAIR manner. For example, the Landmark ESRF EBS (European Synchrotron Radiation Facility Extremely Brilliant Source) uses the ICAT repository from Pandata to store, share and search metadata.[165] Answers to the survey showed that some domains were aware of the metadata related to their fields, while others were unaware and/or need support to understand how to handle metadata.

3.2.4.2. Semantic interoperability and artefacts

Semantic artefacts are chosen based on availability, appropriateness and ease of access. The specialized nature of the field makes it a necessity to have unique repositories that can only be used by those in the field. For example, the Landmark ESRF EBS (European Synchrotron Radiation Facility Extremely Brilliant Source) has two databases - IspyB and TomoDB. However, they are too specific to be applied to other experiments without major modifications.[166]
The formats used for raw data are FITS, ROOT or HDF51.[167,168] The KM3NeT ESFRI uses the ASCII text format as well, and supports mechanisms to support CSV formats for smaller datasets. Some ESFRIs identified domain-specific ontologies of interest. In the Landmark ELT (Extremely Large Telescope, there was an initial goal to use domain-specific ontologies e.g. Foundational Ontologies, Telescope Instrumentation Ontology, OMG SysML Ontology, but lack of resources became a problem.[169]

CIF, the Crystallographic Data Information Syntax and DDC (Dewey Decimal Classification of library contents) were mentioned in the survey. DDC is considered to be insufficient for research topics and was said to be out-of-date. In crystallography, CIF is widely used and there is a committee dedicated to maintaining the standard. The standards are commonly built based on recommendations received from governing bodies e.g International Union of Crystallography (IUCr). Project-driven practices can also become community standards if the projects are large enough to be influential within a specific domain.

3.2.4.3. Identifiers

[165] Götz A. et al. The meta-world of metadata. [web page] ESRF. [cited 21.11.2019] Available from: https://www.esrf.eu/home/UsersAndScience/Publications/Highlights/highlights-2013/et/et8.html
[166] Götz A. et al. The meta-world of metadata. [web page] ESRF. [cited 21.11.2019] Available from: https://www.esrf.eu/home/UsersAndScience/Publications/Highlights/highlights-2013/et/et8.html
[167] KM3NeT Data Management Plan. KM3NeT-INFRADEV GA DELIVERABLE: D4.1. 2017. [report] Available from: https://www.km3net.org/wp-content/uploads/2018/10/D4.1-KM3NeT-Data-Management-Plan.pdf
[168] Baumann TM. & Fanghor H. Control, data acquisition, management and analysis. Presented at SQS Early User Workshop. Schenefeld, 12.02.2018 [presentation] Available from https://www.xfel.eu/sites/sites_custom/site_xfel/content/e35165/e46561/e46889/e69177/e69190/xfel_file69191/20180212_BaumannFangohr_DAQ_eng.pdf
[169] Modeling Guidelines. EELT ICS. [web page] ESO [cited 21.11.2019] Available from: http://www.eso.org/~eeltmgr/ICS/documents/DeveloperGuide/build/html/part-modeling/contents/modguidelines.html

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

The most common persistent identifiers in use are DOIs. KM3NeT is actively involved in the definition of standards via the GEDE-RDA group and Global Neutrino Network (GNN). The goal is to promote persistency, uniqueness and accessibility of data, as the datasets continue to grow.[170]

According to the survey, other PIDs in use are URN, Handle and Database Accession IDs. Some fields also use short URLs for less significant metadata parts. There was a general satisfaction with the currently available PIDs. However it was mentioned that researchers had concerns about GDPR restrictions and were exercising caution by choosing to identify and deposit their data in trusted repositories.

3.2.4.4. RDA CHEMISTRY IG

Chemistry is a fundamental science that is needed, used and applied across various fields e.g. health, pharmaceuticals, materials and energy sciences. However, chemistry data may not be shared across the different disciplines due to limitations such as interoperability issues. In order to promote open sharing and reuse of chemistry research data, an RDA Interest Group on Chemistry Research Data[171] was formed in 2015. The group aims to deliberate over how to promote and improve data management practices within the chemistry community.

3.2.5. Social & Cultural Innovation

'Social and Cultural Innovation' is the title chosen by the European Strategy Forum Research Infrastructures (ESFRI) for the working group dealing with research infrastructures connected to Social Sciences and Humanities.

Social sciences and humanities data cultures have common features, but also many differences. One adjoining feature is that the communities are often in a position to determine what or when something might be used as data. For example, to borrow an example presented by Christine Borgman, to astronomers Galileo's observations are evidence of celestial objects, but to historians those observations may be evidence about the culture at the time (Borgman, 2015).

Scarcity of data has long been a defining feature of humanities research, but with the emergence of digital data resources and computational methods (the so called digital humanities development) the situation is changing dramatically.

Social scientific data commonly describes and originates from contemporary phenomena and sources. The amount of sensitive information in the data is an important defining feature. According to Ron Dekker, rough estimates indicate that 40% of the data need protective measures. These measures can include f.e. anonymization, remote execution, and secured access. In addition to having implications in terms of degrees of openness and accessibility, the sensitive nature can also put limits on data interoperability. It might f.e. be difficult or impossible to connect the data with contextual data, and/or link data using semantics (Dekker, 2019).

[170] KM3NeT Data Management Plan. KM3NeT-INFRADEV GA DELIVERABLE: D4.1. 2017. [report] Available from: https://www.km3net.org/wp-content/uploads/2018/10/D4.1-KM3NeT-Data-Management-Plan.pdf
[171] RDA Interest Group on Chemistry Research Data. [web page] Available from: https://rd-alliance.org/groups/chemistry-research-data-interest-group.html

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

The landscape of the main actors in Europe consists of seven ESFRI roadmap operators: two projects and five landmarks. The projects are European Research Infrastructure for Heritage Science (E-RIHS) and European Holocaust Research Infrastructure (EHRI) and RESILIENCE[172]. The landmarks are Consortium of European Social Science Data Archives (CESSDA ERIC), Common Language Resources and Technology Infrastructure (CLARIN ERIC), Digital Research Infrastructure for the Arts and Humanities (DARIAH ERIC), European Social Survey (ESS ERIC), and Survey on Health, Ageing and Retirement in Europe (SHARE ERIC). Europeana is a DSI (Digital Service Infrastructure), initially the cultural heritage digital library for Europe More about Europeana and its approach to FAIRness in a case study, below.

The Research Data Alliance has many groups that deal with topics relevant to the Social Sciences and Humanities (SSH) domain. These interest and working groups within RDA were recognised as interesting from the point of view of humanities in a report by René van Horik (van Horik, 2019): Digital Practices in History and Ethnography IG, Linguistics Data IG, Mapping the Landscape IG, Social Sciences and Humanities Research Data IG, Ethics and Social Aspects of Data IG, Domain Repositories IG, Empirical Humanities Metadata WG, and Research Data Repository Interoperability WG. A similar exercise was conducted for the social sciences by Ricarda Braukmann (Braukmann, 2018a, 2018b). She categorized the groups into highly relevant and moderately relevant. Many of the groups identified by Braukmann are the same as by Horik, but her list extends wider and therefore copying it here is not useful.

Next we will give an overview of the metadata standards, semantic artefacts, and identifiers in use in the field. This landscape overview relies heavily on the survey (see the chapter on methods). The two case examples (CESSDA and Europeana) are based on desk research. No interview data has been collected from this domain at this time.

Out of the survey respondents, 20 recognised themselves as working with communities from social science and/or humanities. Many of the respondents that identified with the aforementioned domains, identified also with a wide variety of other fields, such as health and medical sciences, environment, and engineering. Also it needs to be taken into consideration that especially humanities, but also social sciences, have a long and close-knit relationship with the field of archiving and archives as institutions. This, to a large extent, is not reflected in the survey results and thus in the analysis, with the exception of the Europeana case example.

3.2.5.1. Metadata standards

Based on the survey, the number of metadata standards in use in the field is quite extensive. For example, one respondent simply wrote that there are “lots”, indicating that there are too many to list. The understanding of what is meant by metadata standard seemed to vary, for example in one instance the FAIR guiding principles were named as a metadata standard. Many of the answers are rather ontologies than metadata standards per se. The framing of the question on what standards are in use in the community seemed to be unclear or confusing to some: one responded with a counter question “by whom?”; another wrote that the standards are domain dependent. These

[172] RESILIENCE [web page] Available from: http://www.resilience-project.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

responses could be translated to indicate how challenging and often artificial defining a scientific/scholarly community is.

The metadata standards that go the most mentions were Dublin Core, Data Documentation Initiative (DDI), Component Metadata Infrastructure (CMDI), and Darwin Core. Among the resources listed there were many that do not respond to a narrow, traditional understanding of social sciences and humanities, f.e. Darwin Core, that deals with biological diversity, or ontologies in the medical domain. This could be taken as an indicator of confusion, or more interestingly, as an indicator of an evolving research landscape, that is becoming more and more multi and transdisciplinary (f.e. global change research, social medicine).

Based on the survey, the biggest metadata related challenge in the domain is not the lack of suitable standards, especially DDI received positive comments as a tailor-made solution, but rather the limited skills of researchers in using them.

Full list of things identified by survey respondents as metadata standards in alphabetical order, number of mentions in brackets if it appeared more than once (n=21): OAI-PMH, BIBFRAME, BIBO, CIDOC-CRM, CMDI (3), Darwin Core (2), DDI (6), Dublin Core (7), EAD, Google Datasets, HGVS Nomenclature, HPO Human Phenotype Ontology, JATS, JSON-LD, KNA (Archeology), META-SHARE, METS, MODS, OpenAIRE, ORDO Orphanet Rare Disease Ontology, RDA, SPAR Ontologies, TEI. In addition, one referred to the SSH recommendations of the SSHOC[173]. All of these do not necessarily normally qualify as metadata standards, such as OAI-PMH, which is more accurately described as a protocol, but we wanted to list them all to point out the varying interpretations, understandings, and perhaps even confusion that exists in the field.

The DARIAH ERIC uses Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for harvesting all metadata from data collections stored in its DARIAH-DE Repository[174], whereas CLARIN's Virtual Language Observatory (VLO) stores indexed metadata in consistent structures and contents to which more than 40 CLARIN centers are sourced on a regular basis from selected OAI-PMH endpoints. The DARIAH Science Gateway service enables semantic interoperability by using a Semantic Search Engine, which holds more than 110 languages for new data discoveries.[175] The problems with metadata quality and semantic interoperability are somewhat revealing for the difficulties in creating really usable services for researchers.

3.2.5.2. Semantic interoperability and artefacts

Twelve of the respondents in this domain responded positively on the use of semantic artefacts, three in the negative. Five respondents didn't know whether semantic artefacts were in use. The negative and unsure answers are more likely to indicate that the respondents are not familiar with the concept, or have a differing understanding of it, from the way it is used f.e. in this report, rather than that they lack any semantic artefacts. It may be more likely that the bodies represented by

[173] D3.1 Report on SSHOC (meta)data interoperability problems. Available from: https://sshopencloud.eu/d31-sshoc-report-sshoc-data-interoperability-problems
[174] DARIAH-DE [web page] Available from: https://repository.de.dariah.eu/publikator/
[175] EOSC Hub D7.2 First Report. 2018. [report] Available from: https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20D7.2%20v1%20Public.pdf

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

these altogether eight respondents have semantic artefacts that they rely on, such as thesauri, vocabulary lists and the like, but that these artefacts are weak in terms of semantic structuration, interoperability, and FAIRness (i.e. machine readability and actionability). This interpretation is further reinforced by the fact that none of these eight elaborated on the situation in the other semantics related survey questions, but instead left the questions unanswered.

DDI controlled vocabularies were the most common semantic artefact recommended with five mentions. This isn't surprising considering how prevalent DDI was as a metadata standard in the responses. DDI CV's and ISO 639-1 (with two mentions) were the only resources with more than one mention. The other semantic artefacts listed were bioportal.bioontology.org, CESSDA Topic Classification, CIDOC-CRM, CLAVAS, DCAT, DARIAH - GND (Gemeinsame Normdatei), ELSST - CLARIN Concept Registry, E-RIHS - AAT (Art and Architecture Thesaurus), ESS' self-defined controlled vocabularies, HASSET[176], HPO (Human Phenotype ontology), ICD, ICD10 (and previous versions), ISO-3166, MeSH, NCIT, Office for National Statistics Classifications, OECD science and technology field classifications, OMIM, GeoNames, ORDO (Orphanet rare diseases ontology), Pactols (archeology), PICO Thesaurus, PROV-O, SNOMED, SPAR Ontologies, TaDIRAH, TGN (Getty Thesaurus of Geographic Names), VIAF (Virtual International Authority File), word2vec semantic vectors, and VoID.

The formats mentioned in the survey are SKOS, OWL, RDF, XML and ShEx. Theoretically, interoperability between semantic artefacts is possible when there are common standards and formats. However, this does not eliminate inconsistencies due to differences in logical or social interpretations as well as varying levels of granularity, licensing, access etc. As pointed out by a survey respondent, "semantically, you can play with European Language Social Science Thesaurus (ELSST) or any other SKOS vocabulary in the Skosmos browser[177]. But that doesn't mean the content is 'interoperable'.

3.2.5.3. Identifiers

In general, the concept and awareness of identifiers seems to be well spread to different communities within the domain. All respondents but one stated that identifiers are used in their community. The outlying respondent was unsure.

For survey respondents, the main motivation for using identifiers was pointing to the object, i.e. citing it or in other ways referring to it. This was mentioned in one way or another in six of the answers. As with metadata standards the lack of suitable identifiers was not considered an issue, unlike wider adoption and improvement of practice.

The survey respondents named five different identifier types in use: DOI (17), URN (8), Handle (6), PURL (4), and ARK (2). Four respondents named also other indicators in use, namely EPIC, HPO Ids, ORCID, Orphanumber, PMC, and database identifiers like arXiv, GenBank ID, PubMed, and Wikidata ID.

[176] Humanities and Social Science Electronic Thesaurus.[web page] UK Data Service. Available from: https://hasset.ukdataservice.ac.uk/
[177] Skosmos [web page] National library of Finland. Available from: http://skosmos.org/

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

3.2.5.4. CASE: CESSDA

Consortium of European Social Science Data Archives CESSDA is one of the five ESFRI landmarks in social sciences and humanities (SSH). It is a European Research Infrastructure Consortium (ERIC) and has sixteen member states and one observer. CESSDA claims to strive for full European coverage eventually.

According to an analysis by Ron Dekker (Dekker, 2019), CESSDA has accomplished the ´F' of FAIR data principles, "is working on the 'A' [...], just started on 'I', and that there is lack of clarity on what should be in 'R'." CESSDA has carried out a self-assessment on its FAIR maturity level using the 27 FAIR Action Plan recommendations presented in the "Turning FAIR into reality" report.

CESSDA has a strategy of working towards FAIRness gradually, starting by focusing on Findability, while working simultaneously with the other principles, but at a more moderate pace. CESSDA's key action towards Findability has been implementing a data catalogue, with metadata on datasets from all their national service providers. The catalogue allows free text search, plus filtering on language, topic, years, country, service provider, and language of data files. Metadata is harvested on a nightly basis. Since the catalogue holds only metadata, there are no privacy or security issues to consider.

Establishing the catalogue required building a metadata harvester that is able to work with the service providers differing systems. Therefore setting up a number of different end-points was necessary.

CESSDA has a persistent identifier policy.[178] It is accompanied by a 'best practices' document.[179]

[178] CESSDA ERIC Persistent Identifier Policy. [web page] Available from: http://multiweb.gesis.org/csaw/#!Detail/cessda-eric/0047
[179] CESSDA ERIC Persistent Identifier Policy Best Practice Guidelines. [web page] Available from: http://multiweb.gesis.org/csaw/#!Detail/cessda-eric/0048

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

3.2.5.5. CASE: Europeana

Europeana is not part of the ESFRI roadmap, but it's one of the European Union's Digital Service Infrastructures (DSI). The main component of a DSI is the core service platform which is a central hub at EU level to which national infrastructures link up and thus create a link between different national infrastructures.[180] From research point of view Europeana is essentially a metadata catalogue that provides access to 57,568,653 artworks, artefacts, books, films and music from European museums, galleries, libraries and archives.[181]

Europeana has developed a quality standard for digital content called the European Publishing Framework. In 2019 a quality standard for metadata was added to the framework[182]. The metadata standard consists of mandatory elements, which are required as a fundamental minimum for all metadata descriptions, and enabling elements. The latter are desirable but optional elements that support functionalities for a specific set of usage scenarios.

Europeana provides its users with a publishing guide that details how to work with the metadata standard.[183] The guidelines encourage the use of language tags to show which language is being used, which facilitates automatic linking and translation processes and allows development of multilingual services. Use of the 'enabling elements' is also encouraged in the metadata. Adding contextual information such as place names, dates and subjects either as metadata elements or as links to contextual vocabularies is also suggested to data publishers.

RightsStatements.org is a joint initiative of Europeana and the Digital Public Library of America (DPLA). It provides standardised international interoperable rights statements to

[180] Connecting Europe Facility (CEF) - Digital Service Infrastructures. EU, 2014. [web page] Available from: https://ec.europa.eu/digital-single-market/en/news/connecting-europe-facility-cef-digital-service-infrastructures
[181] Europeana Collections Portal. [web page]. Available from: https://www.europeana.eu/portal/en
[182] Daley B, Scholz H, Charles V. Developing a metadata standard for digital culture: the story of the Europeana Publishing Framework" [web page] Europeana. [cited 22.11.2019] Available from: https://pro.europeana.eu/post/developing-a-metadata-standard-for-digital-culture-the-story-of-the-europeana-publishing-framework
[183] Europeana Publishing Guide. [web page] Europeana. [cited 22.11.2019] Available from: https://pro.europeana.eu/post/publication-policy

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

the cultural heritage sector. RightsStatements.org currently provides 12 different rights statements that can be used by cultural heritage institutions to communicate the copyright and re-use status of digital objects to the public. The rights statements have been designed with both human users and machine users (such as search engines) in mind and are made available as linked data. Each rights statement is located at a unique URI.[184]

In October 2019 Europeana Research Requirements Task Force released a survey addressed to scholars working in the SSH fields and to researchers working at cultural heritage institutions. It was based on an analysis of state of the art on data management among the Europeana community and ESFRIs. The end-result will be a report on researchers requirements concerning the re-use of digital cultural heritage, followed by recommendations addressed to the Europeana Foundation and the Europeana Network Association.[185] The results of this work will be reviewed in the next versions of this report.

3.2.6. Data, Computing and Digital Research Infrastructures

The Partnership for Advanced Computing in Europe (PRACE[186]) focuses on providing access and guidance for European-wide network of High Performance Computing facilities. PRACE has run workshops for writing Data Management Plans and on technologies on transferring and managing data across network. Apart from this these activities, PRACE sees data solely as a material and results of massive computational efforts. Data is considered to be “Big Data” that is managed as an input to “Data Science” by experts.

4. Conclusions

This report focuses on solutions for semantic interoperability and on persistent identifiers as they are important building blocks of a FAIR ecosystem and framework. We have studied the implementation of semantic interoperability and persistent identifiers in projects and landmarks

[184] About RightsStatements.org. [web page] Available from: https://rightsstatements.org/page/1.0/?language=en
[185] Europeana Research Requirements Task Force [web page] Europeana. [cited 22.11.2019] Available from: https://pro.europeana.eu/project/research-requirements
[186] PRACE. [web page] PRACE. [cited 10.10.2019] Available from: http://www.prace-ri.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

listed by the European Strategy Forum on Research Infrastructures (ESFRI[187]). We have also looked at other stakeholders and activities that are relevant for the implementation of the FAIR principles for research data and what the supporting services and digital infrastructures could do to support and enable FAIR data from a technical point of view. However, the broader implementation of FAIR still requires more specifications and deliberation on context specific solutions.

The two subsequent annual versions of this report will broaden the current scope and follow the development within and around the EOSC projects. We hope to get feedback and comments to help deepen and nuance the presentation and keep it up-to-date. It was not possible to do a thorough analysis of all domains and digital infrastructures within this task, but when other EOSC landscaping activities proceed we can integrate the findings of those into the following reports. This text painted a first outline of solutions that support the FAIR principles. Our main reflections on this work are the following:

1) FAIRness at a more generic level is not ready nor clearly defined. Despite many good efforts, it is very much a work in progress but it will hopefully gain sharper focus once concepts, technologies and implementations mature. For example, at this point FAIR vocabularies, software, and services are largely undefined.
2) The landscape is diverse in all aspects. Differences inside domains are often bigger than differences between domains. Refinement and implementation of the FAIR principles should be driven by research rather than technology to achieve the needed usability and the potentially huge benefits of FAIR data. Standardisation will not solve all problems. The needs of various areas of science have different abilities and needs that will necessitate fine tuning of FAIR evaluation. Community adoption and trust are the decisive factors. For that, practical, easy-to-use implementations are more valuable than precise and high flying, “correct”, and hard to use solutions. Continuous adjustments will be needed as language, technology and science changes.
3) Semantic artefacts are a key element in building interoperability and good quality (meta)data. The maturity and needs are diverse across infrastructures and domains. Shared resources are needed. Management and governance should be ensured. Local data management services need to be involved in both reuse of reference metadata and enabling local modifications. Systematic terminology work and continuous development and curation of knowledge organisation systems is necessary.
4) Crosswalks, mappings and semantic application profiles should be published and registered in machine readable formats.
5) The challenge with PID and data type registries is that they should promote reuse rather than bulk creation of PIDs. To support interoperability, they should be considered semantic artefacts and used mindfully.
6) Reuse of semantic artefacts should be promoted by publishing application profiles. This should happen in machine readable formats in shared registries. Curated registries like the EOSC Hub, FAIRsharing and re3data.org are important resources for promoting implementations of the FAIR data principles.
7) Data citation and machine actionable solutions should be developed in parallel.

[187] European Strategy Forum on Research Infrastructures (ESFRI). [web page] ESFRI. [cited 9.10.2019] Available from: https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/esfri_en

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

8) The most popular, potentially most useful, and most complex approaches on improving FAIRness of data are based on technologies using Linked Data. Their expressiveness and speed of development of new tools and standards is encouraging but at the same time a hindrance to wider adoption. This technology needs to reach a more stable stage and an added level of abstraction that will hide rapidly changing parts from everyday users. At the same time they need to be transparent for the researcher to evaluate.

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. Bibliography

Amstutz, P., Crusoe, M.R., Tijanić, N., Chapman, B., Chilton, J., Heuer, M., Kartashov, A., Leehr, D., Ménager, H., Nedeljkovich, M., Scales, M., Soiland-Reyes, S., Stojanovic, L., 2016. Common Workflow Language, v1.0. https://doi.org/10.6084/m9.figshare.3115156.v2

Aryani, A., 2018. Data Description Registry Interoperability WG: Interlinking Method and Specification of Cross-Platform Discovery. https://doi.org/10.15497/RDA00003

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556

Borgman, C.L., 2015. Big Data, Little Data, No Data: Scholarship in the Networked World. The MIT Press.

Braukmann, R., 2018a. RDA Overview for the Social Sciences - Report. Zenodo. https://doi.org/10.5281/zenodo.1401105

Braukmann, R., 2018b. RDA Overview for the Social Sciences. https://doi.org/10.5281/zenodo.1308776

Chataigner, J., Nowak, C., 2018. French Information System on Water Withdrawals: Challenges of a Data Reuse Project. Biodivers. Inf. Sci. Stand. 2, e25577. https://doi.org/10.3897/biss.2.25577

Dekker, R., 2019. Social Data: CESSDA Best Practices. Data Intell. 220–229. https://doi.org/10.1162/dint_a_00044

Denney, C., Batchelor, C., Bodenreider, O., Cheng, S., Hart, J., Hill, J., Madden, J., Musen, M., Pichler, E., Samwald, M., Szalma, S., Schriml, L., Sedlock, D., Soldatova, L., Sonoda, K., Statham, D., Stenzhorn, H., Whetzel, P.L., Wu, E., Stephens, S., 2009. Creating a Translational Medicine Ontology. Nat. Preced. https://doi.org/10.1038/npre.2009.3686.1

European Commission Expert Group on FAIR Data, 2018. Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data. (Website).

European Union Directorate-General for Informatics, 2017. New European Interoperability Framework - Promoting seamless services and data flows for European public administrations. European Commission. https://doi.org/10.2799/78681

Fabris, E., Kuhn, T., Silvello, G., 2019. A Framework for Citing Nanopublications, in: Digital Libraries for Open Knowledge: 23rd International Conference on Theory and Practice of Digital Libraries, TPDL 2019, Proceedings. Presented at the 23rd International Conference on Theory and Practice of Digital Libraries, TPDL 2019, Springer Verlag, pp. 70–83. https://doi.org/10.1007/978-3-030-30760-8_6

Gregory, K., Cousijn, H., Groth, P., Scharnhorst, A., Wyatt, S., 2019. Understanding Data Search as a Socio-technical Practice. ArXiv180104971 Cs.

Guizzardi, G., 2019. Ontology, Ontologies and the "I" of FAIR. Data Intell. 181–191. https://doi.org/10.1162/dint_a_00040

Jonquet, C., Toulet, A., Dutta, B., Emonet, V., 2018. Harnessing the Power of Unified Metadata in an Ontology Repository: The Case of AgroPortal. J. Data Semant. 7, 191–221. https://doi.org/10.1007/s13740-018-0091-5

Karimova, Y., Castro, J.A., da Silva, J.R., Pereira, N., Ribeiro, C., 2017. Promoting semantic annotation of research data by their creators: a use case with B2NOTE at the end of the RDM workflow, in: Research Conference on Metadata and Semantics Research. Springer, pp. 112–122.

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

Lehväslaiho, H., Parland-von Essen, J., Riungu-Kalliosaari, L., Behnke, C., Laine, H., Staiger, C., Koers, H., LeFranc, Y., 2019. FAIRsFAIR Data of Survey on Semantics and interoperability solutions for D2.1 Report on FAIR requirements for persistence and interoperability. https://doi.org/ 10.5281/zenodo.3518922

McMurry, J.A., Juty, N., Blomberg, N., Burdett, T., Conlin, T., Conte, N., Courtot, M., Deck, J., Dumonter, M., Fellows, D.K., Gonzalez-Beltran, A., Gormanns, P., Grethe, J., Hastings, J., Hériché, J.-K., Hermjakob, H., Ison, J.C., Jimenez, R.C., Jupp, S., Kunze, J., Laibe, C., Novère, N.L., Malone, J., Martin, M.J., McEntyre, J.R., Morris, C., Muilu, J., Müller, W., Rocca-Serra, P., Sansone, S.-A., Sariyar, M., Snoep, J.L., Soiland-Reyes, S., Stanford, N.J., Swainston, N., Washington, N., Williams, A.R., Wimalaratne, S.M., Winfree, L.M., Wolstencroft, K., Goble, C., Mungall, C.J., Haendel, M.A., Parkinson, H., 2017. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. PLOS Biol. 15, e2001414. https://doi.org/10.1371/journal.pbio.2001414

Parland-von Essen, J., Fält, K., Maalick, Z., Alonen, M., Gonzalez, E., 2018. Supporting FAIR data: categorization of research data as a tool in data management. Informaatiotutkimus 37. https://doi.org/10.23978/inf.77419

Rauber, A., Asmi, A., van Uytvanck, D., Proell, S., 2015. Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC). https://doi.org/10.15497/RDA00016

Sansone, S.-A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A.L., Thurston, M., 2019. FAIRsharing as a community approach to standards, repositories and policies. Nat. Biotechnol. 37, 358–367. https://doi.org/10.1038/s41587-019-0080-8

Tomáš Kulhánek, Yann Le Franc, 2019. Service for Digital Annotation of Scientific Data. https://doi.org/10.5281/zenodo.3369156

van Horik, R., 2019. The Research Data Alliance and the Humanities. Zenodo. https://doi.org/10.5281/zenodo.3355145

van Raaij, E.M., 2018. Déjà lu: On the limits of data reuse across multiple publications. J. Purch. Supply Manag. 24, 183–191. https://doi.org/10.1016/j.pursup.2018.06.002

Weigel, T., DiLauro, T., Zastrow, T., 2015. PID Information Types WG final deliverable. https://doi.org/10.15497/FDAA09D5-5ED0-403D-B97A-2675E1EBE786

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M.D., Dumontier, M., Sansone, S.-A., Bonino da Silva Santos, L.O., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., Schultes, E., 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci. Data 6, 174. https://doi.org/10.1038/s41597-019-0184-5

Wittenburg, P., Hellström, M., Zwölf, C.-M., Abroshan, H., Asmi, A., Di Bernardo, G., Couvreur, D., Gaizer, T., Holub, P., Hooft, R., Häggström, I., Kohler, M., Koureas, D., Kuchinke, W., Milanesi,

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

L., Padfield, J., Rosato, A., Staiger, C., van Uytvanck, D., Weigel, T., 2017. Persistent identifiers: Consolidated assertions. Status of November, 2017. Zenodo. https://doi.org/10.5281/zenodo.1116189

Wittenburg, P., Strawn, G., Mons, B., Boninho, L., Schultes, E., 2018. Digital Objects as Drivers towards Convergence in Data Infrastructures. https://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11

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

6. Appendix A. Acronyms and abbreviations

AAT	Art and Architecture Thesaurus
ACTRIS	European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases
ADC	Arctic Data Committee
AnaEE	Infrastructure for Analysis and Experimentation on Ecosystems
ANDS	Australian National Data Service
API	Application Programming Interface
ARK	Archival Resource Key
ASCII	American Standard Code for Information Interchange, a character encoding standard for electronic communication
ASTERICS	Astronomy ESFRI & Research Infrastructure Cluster project
B2FIND	repository metadata discovery service at EUDAT
B2HANDLE	persistent identifier management service for data hosted on EUDAT
B2NOTE	research data annotation service at EUDAT
B2SHARE	EUDAT service to store and publish research data
BARTOC	Basel Register of Thesauri, Ontologies & Classifications
BBMRI-ERIC	European research infrastructure for biobanking
BFO	Basic Formal Ontology
BIBFRAME	Bibliographic Framework, a data model for bibliographic description
BIBO	Bibliographic Ontology
BioMedBridges	joint effort of twelve biomedical sciences research infrastructures on the ESFRI roadmap
BioPortal	repository of biomedical ontologies
BioSchemas	ELIXIR project to add biological types and properties to Schema.org
BNF	Backus–Naur form, a metasyntax notation
CDI	Collaborative Data Infrastructure
CEDAR	The Center for Expanded Data Annotation and Retrieval, https://metadatacenter.org/
CERIC	European Research Infrastructure Consortium for Materials, Biomaterials and Nanotechnology
CERN	the European Organization for Nuclear Research
CESSDA	Consortium of European Social Science Data Archives
CETAF	Consortium of European Taxonomic Facilities
CEVO	Connecting ESFRI projects to EOSC through VO framework
ChEBI	Chemical Entities of Biological Interest
CIDOC-CRM	Conceptual Reference Model of the Documentation Committee of the International Council of Museums, ICOM

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

CIF	Crystallographic Information File, file syntax
CKAN	Comprehensive Knowledge Archive Network, a web-based open-source management system for open data
CLARIN	Common Language Resources and Technology Infrastructure
CLAVAS	CLARIN Vocabulary Service
CMDI	Component MetaData Infrastructure of CLARIN
CNRI	Corporation for National Research Initiatives
COAR	Confederation of Open Access Repositories
CODATA	The Committee on Data for Science and Technology, an interdisciplinary committee of the International Council for Science
CORBEL	Coordinated Research Infrastructures Building Enduring Life-science Services
CORDIS	Community Research and Development Information Service
COST	European Cooperation in Science and Technology, a funding organisation for research and innovation networks
COUNTER	Standard for reporting use of electronic resources in libraries by COAR
CSV	Comma-Separated Values
CTA	Cherenkov Telescope Array, an ESFRI Landmark
CURIE	Compact URI
CV	Controlled Vocabulary
CWL	Common Workflow Language
DADI	Data Access, Discovery and Interoperability for ASTERICS
DARIAH	Digital Research Infrastructure for the Social Sciences and Humanities
DCAT	Data Catalog Vocabulary
DDC	Dewey Decimal Classification of library contents
DDI	Data Documentation Initiative, an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences
DDRI	Data Description Registry Interoperability
DEIMS-DSR	Dynamic Ecological Information Management System - Site and Dataset Registry
DFIG	Data Fabric Interest Group of RDA
DFT IG	Data Foundation and Terminology Interest Group of RDA
DID	Decentralized ID
DiSSCo	Distributed System of Scientific Collections
DKRZ	Deutsches Klimarechenzentrum, German Climate Computing Centre
DMP	Data Management Plan
DO	Digital Object
DOA	Digital Object Architecture
DOI	Digital Object Identifier
DOIP	Digital Object Interface Protocol

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

DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
DSI	Digital Service Infrastructures
DTR	Data Type Registry
EAD	Encoded Archival Description
EASY	Data Archive at DANS
EATRIS	European Infrastructure for Translational Medicine
EBI	European Informatics Infrastructure
EBS	Extremely Brilliant Source of ESFRI
EC	European Commission
ECAS	ENES Climate Analytics Service
ECCSEL	European Carbon Dioxide Capture and Storage Laboratory Infrastructure
ECRIN	European Clinical Research Infrastructure Network
EELT	European Extremely Large Telescope, renamed to ELT in 2017
EGFC	European Group of FAIR Champions
EGI	European Grid Infrastructure
EGO	European Gravitational Observatory
EHRI	European Holocaust Research Infrastructure
EIF	European Interoperability Framework
EIRA	European Interoperability Reference Architecture
EISCAT	European Incoherent Scatter Scientific Association
ELI	Extreme Light Infrastructure
ELIXIR	European Life-sciences Infrastructure for biological Information
ELSST	European Language Social Science Thesaurus
ELT	ESFRI Landmark Extremely Large Telescope
eLTER	European Long-Term Ecosystem and socio-ecological Research Infrastructure
EMBL	European Molecular Biology Laboratory
EMBRC	European Marine Biological Resource Centre
EMFL	European Magnetic Field Laboratory
EMODnet	European Marine Observation and Data Network
EMPHASIS	European Infrastructure for Multi-scale Plant Phenomics and Simulation
EMSO	European Multidisciplinary Seafloor and water column Observatory
ENES	Exchange Network on Exposure Scenarios for chemicals in Europe
ENVRI	ENVironmental Research Infrastructures building FAIR services Accessible for society, Innovation and Research
EOSC	European Open Science Cloud
ePIC	Persistent Identifiers for eResearch
EPOS	European Plate Observing System
ERC	European Research Council

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

ERIC	European Research Infrastructure Consortium
ERINHA	European Research Infrastructure on Highly Pathogenic Agents
E-RIHS	European Research Infrastructure for Heritage Science
ESCAPE	European Open Science of Astronomy & Particle Physics ESFRI research infrastructures
ESFRI	European Strategy Forum on Research Infrastructures
ESIP	Earth Science Information Partners
ESO	European Southern Observatory
ESRF	European Synchrotron Radiation Facility
ESS	European Social Survey
EST	European Solar Telescope
ETag	Entity Tag, a header field of HTTP
EU	European Union
EU-SOLARIS	European Solar Research Infrastructure
EU-OPENSCREEN	European Infrastructure of Open Screening Platforms for Chemical Biology
EUDAT	European Data Infrastructure
EURO-ARGO	European contribution to ARGO project, a global array of autonomous ocean monitoring instruments
Euro-BioImaging	European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences
ExPaNDS	EOSC Photon and Neutron Data Services
FAIR	Findable, Accessible, Interoperable, Reusable; A data management principle
FAIRsFAIR	EU project on understanding FAIR, funder of this report
FinBIF	Finnish Biodiversity Information Facility
FITS	Flexible Image Transport System
FORCE11	a community aiming to improve research communication and e-scholarship
FREYA	A H2020 project aiming to extend the infrastructure for PIDs. Continuation of THOR
GACS	Global Agricultural Concept Space
GBiF	Global Biodiversity Information Facility
GDPR	General Data Protection Regulation
GEANT	pan-European data network for the research and education community
GEDE	Group of European Data Experts in RDA
GEMET	GEneral Multilingual Environmental Thesaurus
GenBank	the NIH genetic sequence database
GeoNames	a global geographical database of freely available place names
GEOSS	Global Earth Observation System of Systems
GND	Gemeinsame Normdatei, Integrated Authority File for catalogue organisation
GNN	Global Neutrino Network

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

GO FAIR	A bottom-up initiative aiming to implement the FAIR data principles
GUISDAP	Grand Unified Incoherent Scatter Design and Analysis Package
HASSET	Humanities and Social Science Electronic Thesaurus
HCLS CG	Semantic Web Health Care and Life Sciences Community Group
HDF	Hierarchical Data Format
HEI	Higher Education Institution
HGVS	Human Genome Variation Society
HL-LHC	High-Luminosity Large Hadron Collider
HPO	Human Phenome Ontology
HTTP	Hypertext Transfer Protocol
IAGOS	In-service Aircraft for a Global Observing System
IBISBA	Industrial Biotechnology Innovation and Synthetic Biology Accelerator
I-ADOPT	InteroperAble Descriptions of Observable Property Terminology Working Group of RDA
ICAT	Initiative for Climate Action Transparency
ICD	International Classification of Diseases
ICOS	Integrated Carbon Observation System
ICS	Instrument Control System
ICSU	International Council of Scientific Unions, name discontinued 2018, now ICS
ICV	Integrity Constraints Validator, Pellet Integrity Constraints, validates RDF with OWL
ID	Identifier
IFDS	Internet of FAIR Data & Services
IFMIF-DONES	International Fusion Materials Irradiation Facility and its DEMO Oriented NEutron Source
IG	Interest Group
IGAD	Interest Group on Agricultural Data of RDA
IGSN	International Geo Sample Number
ILL	Institut Max von Laue-Paul Langevin
INDIGO-Datacloud	INtegrating Distributed data Infrastructures for Global ExplOitation
INFRAFRONTIER	European Research Infrastructure for the development, phenotyping, archiving, and distribution of model mammalian genomes
INSPIRE	Infrastructure for Spatial Information in the European Community
INSTRUCT	pan-European research infrastructure in structural biology
IODE	International Oceanographic Data and Information Exchange
IPFS	InterPlanetary File System
IRP	Identifier/Resolution Protocol
ISA2	Interoperability solutions for public administrations, businesses and citizens in the EU
ISBE	Infrastructure for Systems Biology in Europe

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

ISC	International Science Council
IspyB	Information System for Protein CrystallographY Beamlines
ISO	International Organization for Standardization
IUCr	International Union of Crystallography
JATS	Journal Article Tag Suite
JISC	Joint Information Systems Committee
JIV	Joint Institute for VLBI
JSON	JavaScript Object Notation
KM3NeT	KM3 Neutrino Telescope
KNA	Dutch Archaeology Quality Standard
LifeWatch-ERIC	European Infrastructure Consortium providing e-Science research facilities to scientists seeking to increase our knowledge and deepen our understanding of Biodiversity organisation and Ecosystem functions and services
LOD	Linked Open Data
LOV	Linked Open Vocabularies
M4M	Metadata for Machines
MeSH	Medical Subject Headings
META-SHARE	a sustainable network of repositories of language resources
METROFOOD-RI	Infrastructure for promoting Metrology in Food and Nutrition
METS	Metadata Encoding and Transfer Standard
MIABIS	Minimum Information About Biobank data Sharing
MIBBI	Minimum Information for Biological and Biomedical Investigations
MIRRI	Microbial Resource Research Infrastructure
MOBILISE	Mobilising Data, Experts and Policies in Scientific Collection, project under COST
MOD	Metadata Vocabulary for describing and publishing ontologies
MODS	Metadata Object Description Schema
MOT	Metadata for Ontology Description and Publication Ontology
MPA	Multi-Primary Administrators
MSD	Million Song Database
N2T	Name-to-Thing
NCBO	National Center for Biomedical Ontology
NCIT	National Cancer Institute Thesaurus
NDA	Non-Disclosure Agreement
NetCDF	Network Common Data Form
NI4OS Europe	National Initiatives for Open Science in Europe
NIH	National Institutes of Health of USA
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OBELICS	OBservatory E-environments LInked by common ChallengeS

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

OBO	Open Biological and Biomedical Ontology
OCFL	Oxford Common File Layout
OCI	Open Citation Identifier
OGC	Open Geospatial Consortium
ODIN	DataCite Interoperability Network
ODP	Open Data Plane
OECD	Organisation for Economic Co-operation and Development
OpenAIRE	European Open Science Infrastructure, for open scholarly and scientific communication
OpenCitations	dedicated to open scholarship and the publication of open bibliographic and cita
OpenID	a simple identity layer built on top of the OAuth 2.0 protocol
OMIM	Online Mendelian Inheritance in Man, a catalog of human genes and genetic disorders and traits
OpenAIRE	European Open Science Infrastructure
EU-OPENSCREEN	European Infrastructure of Open Screening Platforms for Chemical Biology
OMG SysML	Object Management Group Systems Modeling Language
ORCID	Open Researcher and Contributor ID
ORDO	Orphanet Rare Disease Ontology
OWL	Web Ontology Language
PANGAEA	Data Publisher for Earth & Environmental Science, https://www.pangaea.de/
PaNOSC	Photon and Neutron Open Science Cloud
PaNRI	Photon and Neutron Research Infrastructure
PICO	Thesaurus del Portale della Cultura Italiana
PID	Persistent identifier
PIDINST	Persistent Identification of Instruments working group of RDA
PMC	PubMed Central
PRACE	Partnership for Advanced Computing in Europe
PROV	Provenance, a Semantic Web standard
PSE	Physical Sciences and Engineering
PubMed	Public MEDLINE-based reference search engine
PubMed Central	a free public repository of beiomedical full-text publications
PURL	Persistent Uniform Resource Locator
R2RML	Relational to RDF Mapping Language
RAiD	Research Activity Identifier
RDA	Research Data Alliance
RDF	Resource Description Framework
RDFA	Resource Description Framework in Attributes
RDFS	Resource Description Framework Schema
RDM	Introduction to Research Data Management

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

RESILIENCE	an EU project to establish resilience as a horizontal theme in adult education
REST	REpresentational State Transfer
ResourceSync	a web synchronization framework
RFC	Request for Comments
RI	Research Infrastructure
RISCAPE	European Research Infrastructures in the International Landscape
RO	Relations Ontology
ROOT	Software toolkit with a machine-independent file format from CERN
SCADM	SCAR Standing Committee on Antarctic Data Management
SeaDataNet	Pan-European Infrastructure for Ocean & Marine Data Management
SensorML	Sensor Model Language
SHACL	Shapes Constraint Language
SHARE	Survey on Health, Ageing and Retirement in Europe
ShEx	Shape Expressions, RDF standard
SKA	Square Kilometre Array
SKOS	Simple Knowledge Organization System
SNOMED	Systematized Nomenclature of Medicine
SOOS	Southern Ocean Observing System
SPAR	Semantic Publishing and Referencing Ontologies
SPARQL	SPARQL Protocol and RDF Query Language
SPIN	SPARQL Inferencing Notation
SPIRAL2	Système de Production d'Ions Radioactifs en Ligne de 2e génération
SQL	Structured Query Language
SSH	Social Sciences and Humanities
SSHOC	Social Sciences & Humanities Open Cloud
STEM	Science, Technology, Engineering and Mathematics
SUSHI	Standardized Usage Statistics Harvesting Initiative
SWORD	Simple Web-service Offering Repository Deposit
SysML	Systems Modeling Language from OMG
TaDIRAH	Taxonomy of Digital Research Activities
TCP/IP	Transmission Control Protocol/Internet Protocol
TeD-T	Term Definition Tool
TEI	Text Encoding Initiative, ontology
TGN	Getty Thesaurus of Geographic Names
THOR	Technical and Human Infrastructure for Open Research
TLS	Transport Layer Security
TMO	Translational Medicine Ontology
TomoDB	Microtomography DataBase

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

UK	United Kingdom
UML	Uniform Modeling Language
UPRI	Unique, Persistent and Resolvable Identifier
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
VIAF	Virtual International Authority File
VISSG	Vocabulary Services IG of RDA
VLO	Virtual Language Observatory
VoID	Vocabulary of Interlinked Datasets
W3C	World Wide Web Consortium
WDS	World Data System
WebID-TLS	WebID-Transport Layer Security, formerly FOAF+SSL
WebSub	an open protocol for distributed publish–subscribe communication on the Internet
WG	Working Group
WindScanner	European WindScanner Facility
WP	Work Package
WUSTL	Washington University in St. Louis
XFEL	European X-Ray Free-Electron Laser
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