



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

D3.5 DESCRIPTION OF FAIRSFair'S TRANSITION SUPPORT PROGRAMME FOR REPOSITORIES

Work Package	WP3 - FAIR data policy and practice
Lead Author (Org)	Marjan Grootveld (DANS)
Contributing Author(s) (Org)	Joy Davidson (DCC), Angus Whyte (DCC), René van Horik (DANS)
Due Date	30.09.2020
Date	29.09.2020
Version	1.0
DOI	https://doi.org/10.5281/zenodo.4058339

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Abstract

FAIRsFAIR is working to better define good practice for repositories through our involvement in certification efforts that enable FAIR data. This document describes a proposed programme of support which will help repositories to adopt these emerging good practices. There is a focus on supporting FAIR data provision, improved handling and integration of metadata, and an increased emphasis on data stewardship to ensure data remains FAIR in the long term.

Versioning and contribution history

Version	Date	Authors/Contributors	Notes
0.1	08.01.2020	Marjan Grootveld	Draft structure.
0.2	21.04.2020	Marjan Grootveld, Patricia Herterich, Josefine Nordling, Joy Davidson, René van Horik	Content added about several WP2, WP3, WP4 topics.
0.3	18.05.2020	Marjan Grootveld	Processed internal feedback.
0.4	27.06.2020	Marjan Grootveld, Patricia Herterich, Joy Davidson, Josefine Nordling, René van Horik, Angus Whyte	Added sections on DMP, metadata, PIDs, and professional development.
0.5	08.07.2020	Marjan Grootveld, Patricia Herterich, Joy Davidson, Josefine Nordling, René van Horik, Angus Whyte, Sarala Wimalaratne, Hervé L'Hours, Anusuriya Devaraju, Claudia Behnke, Hylke Koers, Elizabeth Newbold, Marie-Anne Maurel	Invited internal feedback and comments processed. Draft milestone document finalised for public comment (M3.5).
0.6	31.08.2020	Marjan Grootveld, Joy Davidson	Processed comments on the public milestone draft M3.5. Prepared deliverable for project-internal review.
1.0	29.09.2020	Marjan Grootveld, Angus Whyte, Joy Davidson	Reviewers' comments processed.

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.

Abbreviations and Acronyms

API	Application Programming Interface
CTS	CoreTrustSeal: repository certification scheme
CoreTrustSeal+FAIR	Work being carried out by FAIRsFAIR to extend CoreTrustSeal to better reflect the FAIR principles
DIN 31644	DIN Standards Committee Information and documentation - Criteria for trustworthy digital archives
DMP	Data Management Plan
DOI	Digital Object Identifier
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FDP	FAIR Data Point
ISO 16363	ISO standard for Audit and certification of trustworthy digital repositories
PID	Persistent Identifier
RDA	Research Data Alliance
RPO	Research Performing Organisation
TDR	Trustworthy Digital Repository

Executive Summary

FAIRSF AIR is working to better define good practice for repositories through our involvement in certification efforts that enable FAIR data. D3.5 “Description of transition support programme for repositories” describes a proposed programme of support which will help repositories to adopt these emerging good practices. There is a focus on supporting FAIR data provision, improved handling and integration of metadata, and an increased emphasis on data stewardship to ensure data remains FAIR in the long term.

Table of contents

1. Repositories and FAIR data: a shared journey	6
1.1 Trust, certification and FAIR	7
1.2 Which repositories should prepare for FAIR-enabling certifiable status?	8
2. Recommended actions and FAIRSF AIR support for repositories to become FAIR-enabling	9
2.1 Improve the findability of your repository and its related policies so end users are aware of what you offer	9
2.2 Improve the interoperability potential of your repository content by adding a FAIR Data Point	11
2.3 To support the availability of more FAIR data for reuse, check if your repository is enabling FAIR data	12
2.4 Increase the amount of FAIR data your repository holds through automatic assessment and by improving researchers' awareness of how to make their data FAIR	15
2.5 Support depositors in providing clearer access descriptions by harmonising metadata relating to data accessibility	16
2.6 Improve the visibility of your repository and the data you hold through the use of Persistent Identifiers	17
2.7 Optimise the potential to receive FAIRer data from depositors by supporting machine-actionable Data Management Plan workflows	18
2.8 Invest in professional staff development to ensure that your service meets evolving end-user needs and remains competitive	19
2.9 Support better shared understanding of the costs of keeping data FAIR over time	20
2.10 Benefit from engaging with the European Group of FAIR Champions network	21

1. Repositories and FAIR data: a shared journey

Digital repositories play an important role in enabling FAIR data and keeping it FAIR for the long term throughout changes to technologies and the needs of user communities. They typically ensure that digital objects are technically usable and support interoperability, and are accompanied by persistent identifiers (PIDs), metadata and documentation to support findability, access, and reuse.

The diagrams below (Figure 1) show the components of FAIR digital objects as presented in the “Turning FAIR into Reality” report¹ and how digital objects fit into the wider repository ecosystem.²

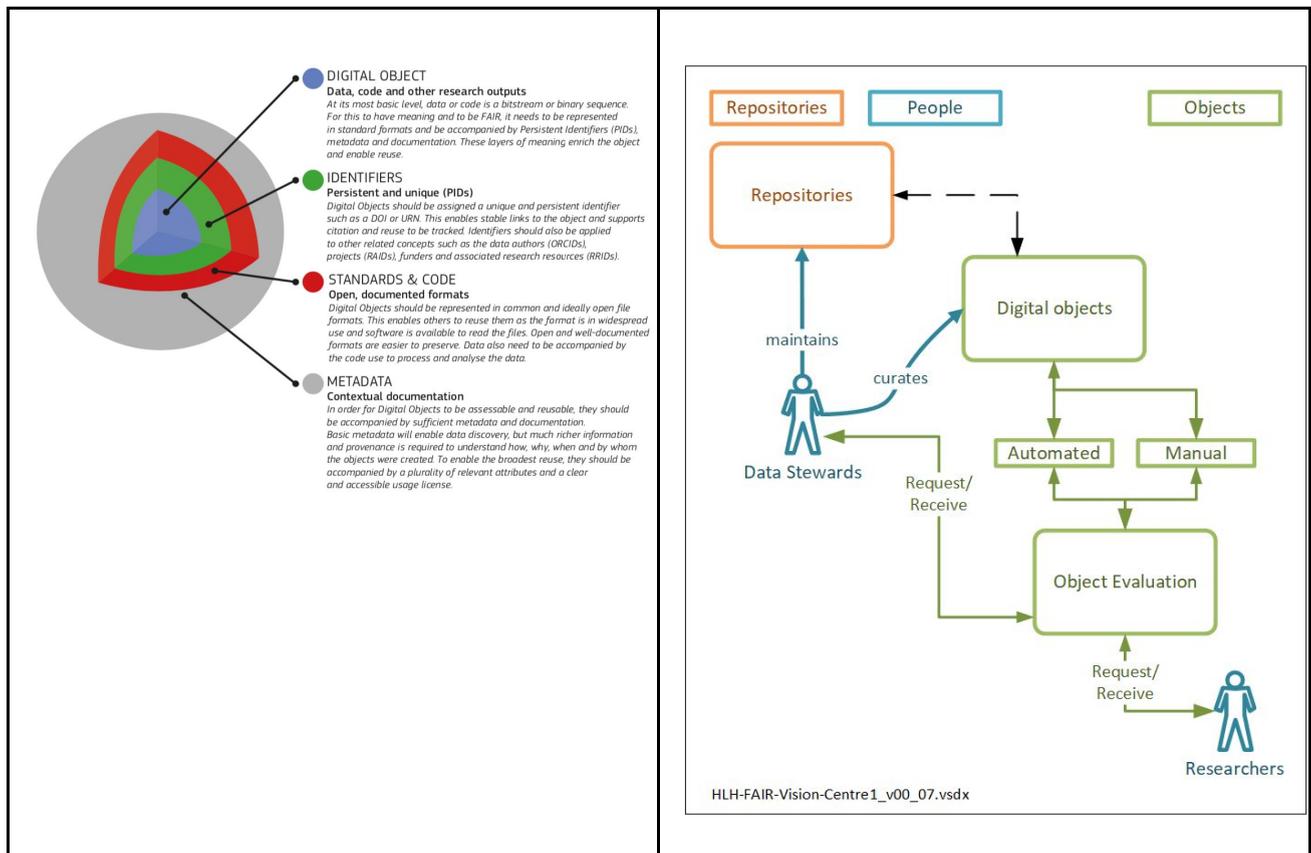


Figure 1. Digital objects and how they fit into the repository ecosystem

The approach underlying this support overview is to offer:

- a rationale for research data repositories to be ‘FAIR-enabling’ i.e. to make the digital objects they hold FAIR digital objects through action by people (data stewards) and processes (evaluation tools).

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. <https://doi.org/10.2777/1524>

² This diagram is part of L’Hours, H., Von Stein, I., 2020. FAIR Ecosystem Components: Vision. FAIRSFAR. <https://doi.org/10.5281/zenodo.3734273>.

- recommendations and supporting action from FAIRSF AIR, that will assist repositories to be FAIR-enabling. This may be as a step to becoming certified as trustworthy, since ideally repositories are both FAIR-enabling and trustworthy. However we appreciate that formal certification (e.g. CoreTrustSeal) may not be not appropriate for all repositories (see below);

Repositories should implement the recommended actions and the associated FAIRSF AIR support to the extent that is feasible and relevant for them.

1.1 Trust, certification and FAIR

Providing repository services that are trustworthy and enable FAIR data is beneficial for future data users, for academia and society at large, but also supports the mission and ambition of repositories to offer a valuable service.

The trustworthiness of digital repositories can be assessed at different levels of rigour and complexity. The European Framework for Audit and Certification proposes three levels of certification: Core (CoreTrustSeal, formerly Data Seal of Approval³ and ICSU-World Data System⁴), Extended (nestor/DIN31664⁵) and Formal (ISO16363⁶). More recently, the RDA community developed the TRUST Principles, which offer guidance for maintaining the trustworthiness of digital repositories, especially those responsible for the stewardship of research data⁷. In FAIRSF AIR, we are focusing on supporting Core Level Certification, embodied by CoreTrustSeal. To ensure better provision of FAIR data, we are working to develop an extension to CoreTrustSeal to better address the FAIR principles. This extension is known as CoreTrustSeal+FAIR⁸ and is a work in progress.

Dillo and De Leeuw⁹ and Donaldson et al.¹⁰ examined the benefits of acquiring certification from the point of view of repositories that achieved the Data Seal of Approval¹¹, one of the two certification schemes from which the current CoreTrustSeal¹² (CTS) certification evolved. There are a number of benefits that can result from achieving certified status. A literature study by Lindlar and Schwab¹³ clusters the benefits of certification in two groups: intrinsic or organisation-internal, and extrinsic, that is, in relation to external parties such as data users. Donaldson et al. report that the repository

³ <https://www.coretrustseal.org/about/history/data-seal-of-approval-synopsis-2008-2018/>

⁴ <https://www.icsu-wds.org/services/certification>

⁵ http://files.d-nb.de/nestor/materialien/nestor_mat_08_eng.pdf

⁶ <https://www.iso.org/standard/56510.html>

⁷ Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020). <https://doi.org/10.1038/s41597-020-0486-7>

⁸ All versions of these emerging requirements can be accessed from <https://doi.org/10.5281/zenodo.3734896>

⁹ Dillo, I., De Leeuw, L., 2015. Ten Years Back, Five Years Forward: The Data Seal of Approval. *International Journal of Digital Curation* 10 (1). <https://doi.org/10.2218/ijdc.v10i1.363>

¹⁰ Donaldson, D.R., Dillo, I., Downs, R., Ramdeen, S., 2017. The Perceived Value of Acquiring Data Seals of Approval. *International Journal of Digital Curation* 12 (1). <https://doi.org/10.2218/ijdc.v12i1.481>

¹¹ <https://datasealofapproval.org>

¹² <https://www.coretrustseal.org/>

¹³ Lindlar, M., Schwab, F. 2018. All that work ... for what? Return on investment for trustworthy archive certification processes – a case study. *iPRES 2018*. <https://doi.org/10.17605/OSF.IO/8A3SC>

representatives they interviewed mentioned increased stakeholder confidence, greater transparency, improvement in repository processes, and increased awareness about digital preservation. These benefits are in addition to the impact of acquiring the certified status on documentation of their workflows and assurance that they were following best practice. Anecdotal evidence also suggests that certified repositories stand a better chance in acquiring external project funding. Indeed, many funding bodies including the European Commission¹⁴ now encourage or mandate grantees to deposit their data and associated metadata with certified repositories which support open access. Through our related efforts in WP4, FAIRSFair will play a key role in the development of global standards for FAIR certification of repositories and the data within them, contributing to those policies and practices that will turn the European Open Science Cloud (EOSC) programme into a functioning infrastructure.

1.2 Which repositories should prepare for FAIR-enabling certifiable status?

The target audience for the FAIRSFair support programme are repositories that wish to make the transition to become more FAIR-enabling, including those that may wish to go on to seek certified status. In principle, all kinds of repositories that deal with research data are expected to make the transition, be they generic or domain-specific, provided by a research-performing organisation, an institute for higher education, a research infrastructure, or as a national service. The focus is on data produced through research, although other kinds of data, such as datasets offered by national libraries or web archives are potentially relevant. Repositories whose purpose is not primarily about making research data accessible are still encouraged to enable FAIR data. They may also potentially meet the requirements of a Trustworthy Digital Repository (TDR) by acquiring the CoreTrustSeal certification. Take Public Sector data for example: although most Public Sector data don't originate in research, they can be valuable resources for research and as such enter the research data life cycle.

Furthermore, we realise that some repositories are more advanced or more FAIR-enabling than others and may already have solutions in place which meet recommendations. This is particularly the case when it comes to long-term preservation¹⁵, which is a core requirement for CTS certification but is not referenced by the FAIR principles. Indeed, CTS-certified repositories may offer valuable insights into some of the issues that should be considered around keeping data FAIR

¹⁴ The EC's "Guidelines on FAIR Data Management in Horizon 2020"

(http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, 2016) encourage use of trusted repositories and this trend looks set to continue into Horizon Europe with some work programmes possibly requiring this as outlined in Box 2: The European Commission leading by example. https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-open-science-cloud-eosc.pdf)

¹⁵ Long-term preservation: "Continued access to digital materials, or at least to the information contained in them, indefinitely," as defined by the Digital Preservation Coalition: <https://www.dpconline.org/handbook/glossary#D> This definition is also used in the CASRAI Research Data Management dictionary (<https://casrai-test.evison.ca/glossary-term/long-term-preservation/>) and the CoreTrustSeal glossary 2020-2022 (<https://doi.org/10.5281/zenodo.3632563>).

in response to changes over time. We encourage them to share any lessons they have learned with FAIRSF AIR and with the wider repository community.

While repository service managers are clearly knowledgeable about the repositories they run, they are not necessarily familiar with the FAIR principles or certification processes that help to assess the repository's trustworthiness. Drawing on parallel activities across the FAIRSF AIR project¹⁶, this report suggests practical actions that can help repositories progress towards FAIR-compliant status and describes what support might be needed to facilitate action. Moving towards Trustworthy Data Repository (TDR) status and achieving FAIR data provision won't happen overnight and should be considered a journey. This short document intends to help repository service managers start the journey towards becoming a FAIR-enabling digital repository that is better prepared to seek certified status.

2. Recommended actions and FAIRSF AIR support for repositories to become FAIR-enabling

In this section, we outline a series of practical actions that repositories can take to improve enabling FAIR data. For each action, we provide some contextual background along with a description of the support that FAIRSF AIR will provide.

2.1 Improve the findability of your repository and its related policies so end users are aware of what you offer

One of the best things you can do to support the emergence of a FAIR ecosystem is to ensure that your repository is visible and can be easily found by researchers seeking to deposit or access FAIR digital objects. One way to do this is to add your repository service to registry services such as re3data¹⁷ offered by DataCite¹⁸ which holds information on more than 2500 research data repositories from across the globe; or to FAIRsharing¹⁹, an RDA-endorsed²⁰ resource used by publishers, funders and stakeholders that connects repository records to the data and metadata standards that the repository implements. Thanks to the many filters these registries offer, users can search for repositories that meet their specific needs. Through the FAIRSF AIR project, DataCite is developing a filter in their Repository Finder tool²¹ to enable searching for repositories holding FAIR data content. The Repository Finder tool has been extended to query the re3data registry for

¹⁶ In particular, this document draws on parallel activity in WP2 "FAIR Practices: Semantics, Interoperability, and Services", WP3 "FAIR Data Policy and Practice", WP4 "FAIR Certification", and WP6 "FAIR Competence Centre". More information and project deliverables are available from the project website <https://www.fairsfair.eu/>

¹⁷ <https://www.re3data.org/>

¹⁸ <https://datacite.org/>

¹⁹ https://fairsharing.org/communities#activity_2

²⁰ "RDA Recommendations are the official, endorsed results of RDA [Research Data Alliance] and considered our "flagship" outputs." <https://rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs>

²¹ <https://repositoryfinder.datacite.org/>

repositories relevant to FAIRSF AIR on e.g. Open Access to data (and where needed Restricted Access to sensitive data as well), adoption of various kinds of Persistent Identifiers to the data, and certification against various certification schemes. Like the FAIR principles the filter is discipline-independent. The filter is available for testing and feedback is welcomed.

In addition to making your repository findable, you should also aim to make its related policies visible - both to humans and machines. Those policies should make clear what content your repository will accept, what formats you support and whether you are willing to accept sensitive content that may require added security measures to be accessed. By ensuring your policies are easy to find, potential depositors can make informed choices about where they store their data. By making your policies findable and readable by machines too, you are also helping to support increased automation of some research data management processes (e.g., machine-actionable data management plans). Furthermore, by ensuring that portals and aggregators can harvest your metadata, you make your data discoverable. Similarly, you can support search engines to crawl and index your website, for instance by a Google Sitemap²². And by aligning with the OpenAIRE Guidelines²³, repositories can make their research data - at item level - discoverable. These guidelines have been developed by the community and are based on an inclusive approach to adopting established standards.

Recommended actions:

- Register your repository with a registry such as re3data or FAIRsharing and keep the information up to date.
- Refine and/or develop clear policies relating to the repository service. Policies should be regularly reviewed.
- Ensure that policies are visible to both humans and machines so that they can be put into action.
- Optimise your metadata and your website for metadata harvesters and search engines, respectively. Ensure that portals and aggregators can harvest your metadata; support search engines to crawl and index your website, for instance by a Google Sitemap²⁴.
- Align your repositories with the OpenAIRE Guidelines.

What FAIRSF AIR will provide:

- Guidance on ways to improve the findability of your repository's FAIR content.
- Further integration of the Repository Finder tool to link repository information within the DataCite ecosystem (e.g., with links to DataCite DOIs, institutional identifiers).
- Guidance on aligning policies with FAIR and on making the policies themselves FAIR.

²² <https://support.google.com/webmasters/answer/156184>

²³ <https://guidelines.openaire.eu/en/latest/data/index.html>

²⁴ <https://support.google.com/webmasters/answer/156184>

2.2 Improve the interoperability potential of your repository content by adding a FAIR Data Point

Repositories vary in the way that they capture metadata (e.g., information about the deposit such as “author”, “creator”, and “depositor”), and almost by definition keywords are semantically ambiguous. This is tricky for users who want to query multiple repositories to find relevant data. This is an interoperability issue: repositories not fully “speaking the same language”. Metadata aggregators remedy this by harmonising metadata, however, this is obviously limited to just those repositories whose metadata they harvest.

FAIRSFair has released a list of FAIR data repository features²⁵ to support the emergence of an innovative FAIR data infrastructure, which involves the introduction of an additional interoperability layer and APIs built up around the principles of FAIR practice. A FAIR Data Point²⁶ (FDP) allows a repository to provide metadata in a FAIR-enabling manner, and this helps both humans and machines to discover the data. FDPs have been implemented by several services, offering good examples of well-structured metadata mappings. These include B2SHARE²⁷ and EOSC-Pillar.²⁸ The software²⁹ uses a REST API for FAIR metadata creation, storage and provision and allows a two-way exploitation of the digital objects. The digital objects are not restricted to datasets only, rather the metadata can also expose ontologies, repositories, web resources etc. The ultimate goal is to enable interoperability and use of digital objects in different repositories, but also to establish a common understanding of the structure and composition of metadata schemas.³⁰

FAIRSFair cooperates with six ‘developer repositories’ and six ‘tester repositories’.³¹ The first step is to build a prototype with DCAT 2³² in the form of an interoperability layer, which is the subsequent version of the FAIR Data Point. FAIRSFair will provide a reference implementation. The repositories can choose to implement their own reference implementations in their repositories, or use the prototype developed within the project. FAIRSFair has allocated a limited amount of resources to assist in mapping the metadata schemas provided by developer repositories. Tester repositories will get access to a “sandbox” reference implementation to find gaps in the metadata. The first

²⁵ Behnke, C., Bonino, L., Coen, G., Le Franc, Y., Parland-von Essen, J., Riungu-Kalliosaari, L., Staiger, C., 2020. Set of FAIR data repositories features. FAIRSFair D2.3. <https://doi.org/10.5281/zenodo.3631528>

²⁶ Video introducing the FAIR Data Point: https://www.youtube.com/watch?v=PtS_ek7BXSA

²⁷ Moreira, J., Bonino, L., Pires, L., Van Sinderen, M., Henning, P. 2019. Towards Findable, Accessible, Interoperable and Reusable (FAIR) Data Repositories: Improving a Data Repository to Behave as a FAIR Data Point. <https://doi.org/10.18617/liinc.v15i2.4817> and <https://github.com/jonimoreira/B2SHARE-FAIR/wiki>

²⁸ <https://www.eosc-pillar.eu/establishing-fair-data-services>

²⁹ <https://github.com/FAIRDataTeam/FAIRDataPoint>

³⁰ FAIR Data Point specification: <https://github.com/FAIRDataTeam/FAIRDataPoint-Spec>

³¹ The developer and tester repositories are listed here:

<https://www.fairsfair.eu/application-results-open-call-data-repositories>

³² <https://www.w3.org/TR/vocab-dcat-2/>

reference implementation of the repository interoperability features is planned to be ready in February 2021 and the second implementation one year later.

Recommended actions:

- Use agreed vocabularies, terminologies and metadata standards, as listed by FAIRsharing³³.
- Familiarise yourself with the list of FAIR data features described as necessary for enabling an innovative FAIR data infrastructure and consider how well your repository currently supports these.

What FAIRsFAIR will provide:

- Guidance for repositories on how to assess their readiness to support FAIR data features and tips on how to improve their FAIR-enabling practices
- A reference implementation of the repository interoperability layer in an open source format
- A sandbox environment for testing the reference implementation
- Share lessons learned about setting up a FAIR Data Point

2.3 To support the availability of more FAIR data for reuse, check if your repository is enabling FAIR data

Starting from the notion that *data* should be FAIR, the question arises to what extent and how data *services* such as repositories can support the provision of FAIR data. For data services in general, Koers et al. make recommendations that are based on a series of community workshops.³⁴ For repositories in particular, CoreTrustSeal (CTS) certification is highly relevant because the CTS requirements align well with and complement the FAIR data principles. While FAIR and CTS tackle the issue from different perspectives - i.e., FAIR at the data level and CTS at the repository level - they are driving towards the same objective, which is to make data reusable. The mapping in figure 2 below illustrates the alignment between CoreTrustSeal requirements and the FAIR principles³⁵, where “R” notes a CTS requirement.

³³ <https://fairsharing.org/standards>

³⁴ Koers, H., Bangert, D., Hermans, E., Horik, R. van, Jong, M. de, & Mokrane, M. (2020). Recommendations for Services in a FAIR Data Ecosystem. *Patterns*, <https://doi.org/10.1016/j.patter.2020.100058>

³⁵ L’Hours, H., Von Stein, I., Mokrane, M., Devaraju, A., Huigen, F., Davidson, J., De Vries, J., Herterich, P., 2020. CoreTrustSeal plus FAIR Overview. FAIRsFAIR. <https://doi.org/10.5281/zenodo.3862616> p. 16

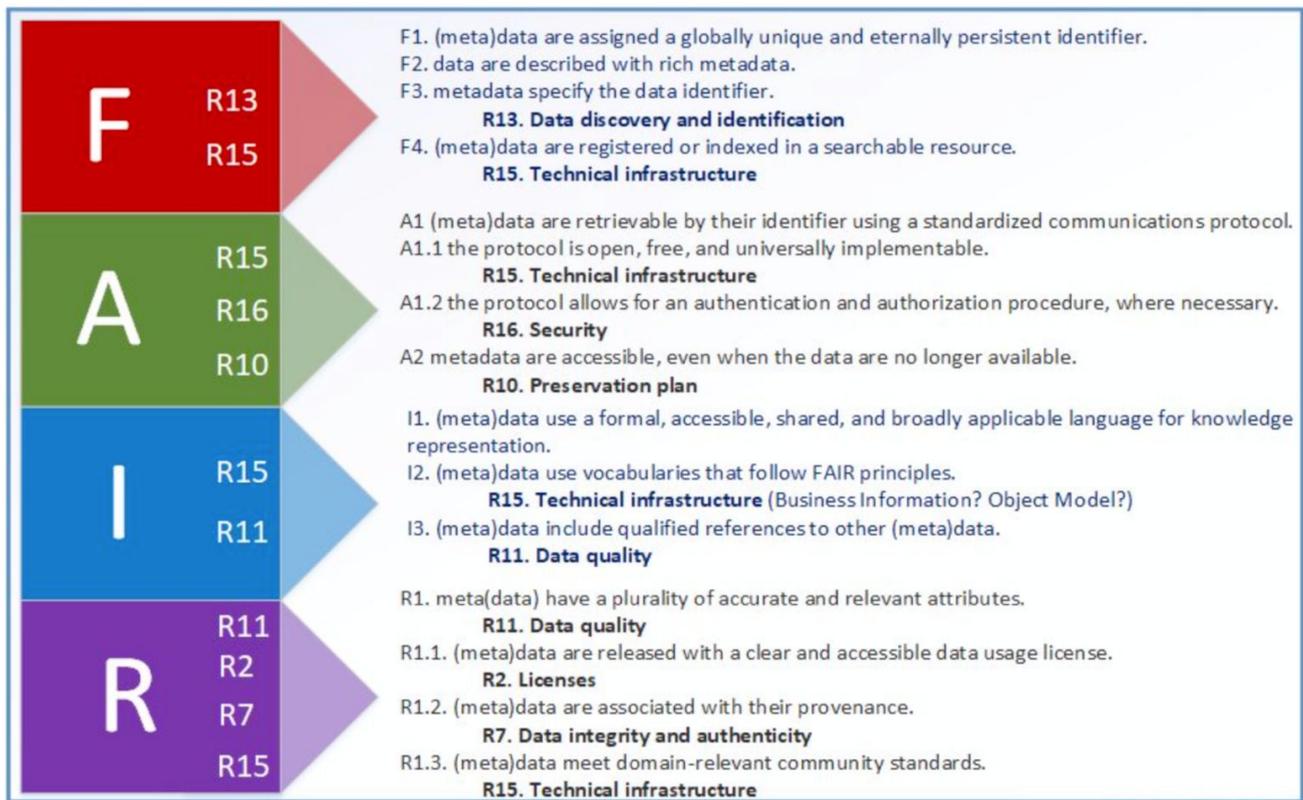


Figure 2. CoreTrustSeal requirements mapped to FAIR Principles - Findable, Accessible, Interoperable, and Reusable

While there are several areas of overlap, one shouldn't conflate a FAIR-enabling repository with a Trustworthy Digital Repository. For example, TDRs are expected to have a mission to ensure long-term preservation of content, however, the FAIR principles do not specify any length of retention period. To meet both ambitions - that is, to work towards CoreTrustSeal+FAIR - a repository needs to maintain its FAIR-enabling support over time.

FAIRSF AIR started work on a maturity evaluation approach to align the characteristics of FAIR digital objects with the repositories that enable FAIRness, through the CTS requirements. The work includes the FAIR indicators developed by the RDA FAIR Data Maturity Model Working Group³⁶. This approach will lead to recommendations for future iterations of the CTS requirements. An updated version of "CoreTrustSeal plus FAIR overview" is available.³⁷ Work on CoreTrustSeal+FAIR and on FAIR assessment will be used to inform recommendations on object and repository metadata.

³⁶ <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

³⁷ L'Hours, H., Von Stein, I., Mokrane, M., Devaraju, A., Huigen, F., Davidson, J., De Vries, J., Herterich, P., 2020. CoreTrustSeal plus FAIR Overview. FAIRSF AIR. <https://doi.org/10.5281/zenodo.3862616>

All interested data repositories are encouraged to consider certification against the CoreTrustSeal's requirements for Trustworthy Digital Repositories, see the website³⁸ and the short video.³⁹ Alternative to the CoreTrustSeal requirements, which for some may not (yet) be attainable, you may find inspiration in the TRUST principles for digital repositories, a framework focusing on Transparency, Responsibility, User focus, Sustainability and Technology⁴⁰. Furthermore, although targeting researchers and research funders rather than repositories, Science Europe's criteria and guidance for the selection of trustworthy repositories⁴¹ may be useful (note that the term "trustworthy" doesn't imply formal certification there, however, "it is strongly recommended that repositories that have not yet been certified seek certification" (p.12)). Even if your repository is not currently seeking certified status, you should still work towards ensuring that you are working to meet recognised good practices. A good way to do this is to develop a roadmap outlining which areas of good practice are priorities for your repository and outlining how you will implement these.

Recommended actions:

- Familiarise yourself with the mapping of FAIR principles and CoreTrustSeal, to better understand how these relate to each other.
- Assess your repository's current ability to support FAIR data provision.
- Develop a roadmap outlining areas for improvement that is aligned with recognised good practice.

What FAIRsFAIR will provide:

- In consultation with repositories, FAIRsFAIR will produce recommendations for a "CoreTrustSeal plus FAIR" specification.
- Guidance and support to help repositories align with the FAIR principles and to be better prepared for releasing CoreTrustSeal plus FAIR-enabling status.
- Materials and guidance on developing a roadmap for repositories aspiring to CoreTrustSeal certification.⁴²

³⁸ <https://www.coretrustseal.org/>

³⁹ Walker, E., Ditsen de Jong, Y., De Bruin, K. and Chatzipoufli, N.D., 2019. CoreTrustSeal - Awareness-sub. This material is made by the authors as part of a cooperation project between the Master Digital Design of the Amsterdam University of Applied Sciences and DANS (Data Archiving and Networked Services). October 2019. Video licensed under a Creative Commons Attribution 4.0 International License. Retrieved from:

https://dans.knaw.nl/en/current/news/coretrustseal-in-2-minutes?set_language=en

⁴⁰ Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020).

<https://doi.org/10.1038/s41597-020-0486-7>

⁴¹ Science Europe, 2018. Practical guide to the international alignment of research data management. D/2018/13.324/4.

https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

⁴² Von Stein, I., Huigen, F., Mokrane, M., 'Hours, H., Herterich, P., Devaraju, A., Rouchon, O., 2020. Certification + FAIR Support Workshop for Data Repositories. FAIRsFAIR. <https://doi.org/10.5281/zenodo.3754292>

2.4 Increase the amount of FAIR data your repository holds through automatic assessment and by improving researchers' awareness of how to make their data FAIR

The FAIR principles are high-level guidelines and leave the users to decide on their implementation. This leads to a range of sometimes ambivalent or contradictory interpretations, raising the need to define systematic measurements of data FAIRness. FAIRsFAIR follows a use-case driven iterative approach to develop a set of minimum viable metrics to assess FAIRness of research data. The current metrics⁴³ have been updated based on focus group evaluation and the final version of the RDA FAIR Data Maturity Model⁴⁴, whereas earlier versions already took other assessment tools into account. In addition to the metrics, the project partners have explored several FAIR data assessment scenarios which are relevant in different stages of the data life cycle. The team shortlisted two use cases to implement metrics tools and address FAIR assessment scenarios.⁴⁵ The first planned use case focuses on raising awareness of researchers who can self-assess their level of familiarity with FAIR before depositing their data into a repository. A prototype version of the tool (named FAIR-Aware) is now available and will be iteratively developed based on feedback⁴⁶ The second planned use case focuses on the automated assessment of data objects deposited and published in selected data repositories. For this use case, PANGAEA developed a web service (named F-UJI)⁴⁷ to demonstrate programmatic assessment of datasets in the repositories, based on generally applicable data/metadata characteristics until domain/community-driven criteria have been agreed. To encourage wider application of the service developed, the source code of the service is made available through github under a public licence⁴⁸.

Recommended actions:

- Become familiar with the FAIRsFAIR assessment metrics and consider how these might relate and/or be supported by your repository processes and workflows.
- Test the pilot version of the self-assessment tool and automated service and provide feedback.
- Share use cases relating to FAIR data assessment via the FAIRdata Forum⁴⁹.

What FAIRsFAIR will provide:

- Metrics to assess the FAIRness of datasets, addressing some of the use cases.

⁴³ Devaraju, A., Mokrane, M., Herterich, P., De Vries, J., Davidson, J., Huber, R., and Cepinskas, L., 2020. "FAIRsFAIR Data Objects Assessment Metrics". FAIRsFAIR. <https://doi.org/10.5281/ZENODO.3775793>

⁴⁴ RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. <https://doi.org/10.15497/RDA00050>

⁴⁵ Devaraju, A., Herterich, P., 2020. Draft Recommendations on Requirements for Fair Datasets in Certified Repositories. FAIRsFAIR D4.1. <https://doi.org/10.5281/zenodo.3678716>

⁴⁶ <https://fairaware.dans.knaw.nl/>

⁴⁷ <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

⁴⁸ <https://github.com/pangaea-data-publisher/fuji>

⁴⁹ <https://fairdataforum.org/>

- An automated tool for assessing the FAIRness of data in repositories.
- Access to collected use cases relating to the assessment of FAIRness of datasets.

2.5 Support depositors in providing clearer access descriptions by harmonising metadata relating to data accessibility

Data held in FAIR-enabling repositories should be findable. This does not mean that all data held in such repositories must also be publicly accessible without restriction though. Commonly accepted exceptions to openly sharing research data include data that can be related to individuals (person-related data), commercially sensitive data, and security-related data. Data that cannot be shared can still be FAIR as long as the access conditions and reasons for not sharing are made explicit. Recent research by Horton et al. found that there are numerous access categories in use across repositories and that there is potential to reduce these to support harmonisation.⁵⁰ In other words, this is not about restrictions that depositors or funders may place on access to the data, but on the different “flavours” of “open”, “restricted”, “controlled” et cetera that repositories offer. In addition to harmonising the access categories themselves, there is also a need to define a list of standard exceptions to sharing data that could be added to metadata schemas to provide additional context.⁵¹ Standardised descriptions for data sharing exceptions in metadata schemas could help pave the way for automated processing over the research lifecycle from the data management planning stage through to ingest, access and legitimate reuse.

In some cases, data may only be kept for a finite period and then for legal reasons must be destroyed. In such cases, repositories should develop tombstone metadata records that are maintained even when data is no longer available. Repositories should also have an explicit data deletion policy.⁵² Like other metadata, the tombstone record should be interpretable by both humans and machines. For more information see DataCite’s “Best practices for tombstone pages”.⁵³

Recommended actions:

- Consider harmonising the data access categories in use at your repository.
- Contribute to defining standard exceptions to data sharing in metadata schemas.
- Publish a data deletion policy that includes reference to maintaining tombstone records.

What FAIRSFAR will provide:

⁵⁰ Horton, L., Perry, A., Bishop, L., 2020. Open where possible, closed if necessary: reforming access categories for social science data archives. <http://doi.org/10.5281/zenodo.3670943>

⁵¹ As recommended in Davidson, J., Grootveld, M., Whyte, A., Herterich, P., Engelhardt, C., Proudman, V., Stoy, L., 2020. Policy Enhancement Recommendations. FAIRSFAR D3.3. <https://doi.org/10.5281/zenodo.3686901>

⁵² As outlined on page 11 of the FAIRSFAR report about data repository features. Behnke, C., Bonino, L., Coen, G., Le Franc, Y., Parland-von Essen, J., Riungu-Kalliosaari, L., Staiger, C., 2020. Set of FAIR data repositories features. FAIRSFAR D2.3. <https://doi.org/10.5281/zenodo.3631528>

⁵³ <https://support.datacite.org/docs/tombstone-pages>

- Examples of emerging good practice in relation to harmonising data access categories.
- Examples of exceptions to data sharing that may serve as the basis for ongoing harmonisation efforts.
- Promote existing guidance on how to develop tombstone records.

2.6 Improve the visibility of your repository and the data you hold through the use of Persistent Identifiers

Persistent Identifiers (PIDs) are core to FAIR Digital Objects (see Figure 1). The EOSC FAIR Working Group and EOSC Architecture Working group have recently published the “Second draft Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)”.⁵⁴ The PID Policy foresees “a future where PIDs can be used as the preferred method of referring to its assigned entity, where appropriate, alongside human-readable means e.g. the common name.” Furthermore, it states that “PID Service Providers and repositories must have clear policies and guidelines on how to manage versioning in case the FAIR Digital Object or entity changes”.

Your repository should be able to assign globally unique PIDs to the content that it ingests and provide metadata fields that support linking between FAIR digital objects - both those held within your repository and also externally. PIDs are a fundamental building block for citing research outputs, which in turn is essential for transparent research and giving credit where credit is due. While the academic world is used to using a PID to cite a publication, citing data and other digital outputs such as software are not yet common practice. Make it easy for end users to correctly cite the data they find in your repository by providing a recommended citation along with mechanisms to export the citation in various formats.

The use of identifiers for data is essential for supporting FAIR data. However, there are a number of additional identifiers that should also be employed to provide better provenance information and context about the data itself. These include identifiers for other actors in the research ecosystem such as funders, researchers, organisations and entities such as projects. In support of transitioning towards enabling FAIR data, repositories may wish to acquire a unique digital identifier to support disambiguation and machine readability. In addition to DOIs for data and ORCIDs for researchers, you may consider supporting unique identifiers for organisations such as those provided by the Research Organisation Registry (ROR)⁵⁵ - a community-led project to develop an open, sustainable, usable, and unique identifier for every research organisation in the world.

⁵⁴ Hellström, M., Heughebaert, A., Kotarski, R., Manghi, P., Matthews, B., Ritz, R., Conrad, A., Weigel, T., Wittenburg, P., 2020. Second draft Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC). <https://doi.org/10.5281/zenodo.3780423> The final policy will be delivered to the EOSC Governance Board in October 2020.

⁵⁵ <https://ror.org/about/>

The FREYA Project has compiled short guides to help with choosing persistent identifiers for the entities Publications, Datasets, People, Organisations and Software. These guides are designed to provide a starting point for anyone thinking about using persistent identifiers in their systems.⁵⁶

Emerging PID Graph or Research Graph technology utilises PIDs to show how various actors and other digital objects relate to each other. The richer contextual information this provides can help to inspire trust in the research output, which in turn may lead to more re-use. More information about this is provided by the FREYA project⁵⁷ and OpenAIRE.⁵⁸

Recommended actions:

- Implement PIDs for different entities, such as scholarly output, researchers, organisations and research funders.
- Provide a recommended citation format for data held within your repository.
- Familiarise yourself with PID graphs.

What FAIRsFAIR will provide:

- Working with research communities and other actors, we will develop guidance to help stimulate a culture of data citation.
- Guidance on how to support the use of a range of unique identifier systems within your repository to support a FAIR ecosystem.

2.7 Optimise the potential to receive FAIRer data from depositors by supporting machine-actionable Data Management Plan workflows

Complementary to project plans, researchers create Data Management Plans (DMPs), in which they describe how they will deal with the data generated, captured, re-used, processed, analysed and shared in their project. A DMP should make clear which repository will be used and, particularly in cases where institutional repositories will be used, a DMP can be a valuable communication tool between the researchers and the repository in relation to storage and metadata requirements, preferred file formats, and licensing. Making DMPs ‘machine-actionable’ means making their content findable and accessible, exchanging that content with other systems in standardised, interoperable ways, and potentially reusing that content. A standard for exchanging DMP content has recently been developed by an RDA working group⁵⁹, with early adoption by some DMP platforms. By supporting the RDA standard, repositories can potentially be better informed about the choices DMP authors are making from the earliest stages of their research. By taking a more

⁵⁶ The first version of the guides to choosing Persistent Identifiers can be found at: <https://zenodo.org/record/3862656>

⁵⁷ <https://www.project-freya.eu/en/pid-graph/the-pid-graph>

⁵⁸ Manghi, P., Bardi, A., Atzori, C., Baglioni, M., Manola, N., Schirrwagen, J., Principe, P., 2020. The OpenAIRE Research Graph Data Model (Version 1.3). <http://doi.org/10.5281/zenodo.2643199>

⁵⁹ Miksa, T., Walk, P., Neish, P., 2019. RDA DMP Common Standard for Machine-actionable Data Management Plans. <https://doi.org/10.15497/rda00039>

active role in ensuring that the data they eventually receive is FAIRer, repositories can reduce the need to FAIRify data retrospectively, which is generally more resource intensive.

Recommended actions:

- Familiarise yourself with the RDA Common Standard to exchange DMP content.⁶⁰
- Consider how you might provide information about your repository's policies and practices to help people create DMPs.
- Consider how your ingest and data management processes could benefit from information contained in DMPs, and provide researchers with repository-related information for their DMP (e.g. about metadata standards, PIDs, and sustainable and where possible open file formats)
- Consider implementing software support for the Common Standard in your data ingest workflows, e.g. using APIs provided by DMP platforms

What FAIRsFAIR will provide:

- Examples and use cases for implementing machine-actionable DMPs, in collaboration with the RDA Exposing DMP working group.⁶⁰

2.8 Invest in professional staff development to ensure that your service meets evolving end-user needs and remains competitive

Data stewards are staff from research communities and research libraries who support researchers and research organisations in managing data throughout the data life cycle. As can be seen in Figure 1, data stewards are an essential liaison between researchers who produce data and repositories who provide long-term access to them. While data stewards are often based in research performing organisations, their expertise regarding data curation and preparing data for sharing also fits alongside the front office activities of repositories - “speaking the same language” is important for institutional, discipline-specific, and generic repositories. Therefore repositories should be familiar with current data stewardship training, or even get involved in it, for the benefit of their own staff (as participants) and/or in the training (as tutors).⁶¹ FAIRsFAIR is building on the successful CODATA/Research Data Alliance schools model⁶², which provides early career researchers with foundational data science skills in a two-week curriculum. FAIRsFAIR seeks collaboration with other EOSC-related projects to adopt the curriculum, propagating the skills by “training the trainers”, and supplying franchised modules which can be tailored for a particular

⁶⁰ <https://rd-alliance.org/groups/exposing-data-management-plans-wg>

⁶¹ <https://www.fairsfair.eu/events/training>

⁶² <https://codata-rda-datascienceschools.github.io/DataSteward/>

community. Training trainers has a multiplier effect and helps to build the capacity that is needed to deal with the shortage in FAIR data support professions.⁶³

Recommended actions:

- Explore FAIRsFAIR training materials.
- Consider how to make your training and learning material FAIR.⁶⁴
- Share your FAIR-related training and learning materials with peers.

What FAIRsFAIR will provide:

- A networking platform to support peer to peer knowledge exchange via the FAIR Competence Centre.⁶⁵
- Access to information on emerging standards standards for the EOSC on training and skills development including those being defined by the RDA, GO-FAIR and other fora.
- Training for data stewards and trainers

2.9 Support better shared understanding of the costs of keeping data FAIR over time

An increasing number of research funding bodies will support justified costs associated with making data FAIR. While the Research-Performing Organisation can help its researchers to estimate costs associated with the active stage of research, the costs associated with long-term curation must be articulated by the repository. There is a possible tension between what the repository can currently offer with the resources it has and the need for value-added services to FAIRify data upon ingest. Repositories must review what level of service they can provide under their current business model and consider whether additional income streams will be required to support FAIR data provision over time. Costs are also strongly related to the amount of curation a repository performs over time, and “naïve” researchers, RPOs or funders may fail to understand why repository X charges so much more than repository Y. Repositories should make clear the levels of basic service being offered and make clear any costs that may be involved in providing value-added services. The online Curation Cost Exchange supports understanding and comparing digital curation costs⁶⁶. Developing a clear catalogue of costs for services will help to ensure that researchers - and more likely RPOs - can factor these into grant proposals as eligible costs.

⁶³ More about developing professional support by data stewards and research software engineers in Molloy, L., Nordling, J., Grootveld, M., van Horik, R., Whyte, A., Davidson, J., Herterich, P., Martin, I., Méndez, E., Principe, P., Vieira, A., Asmi, A., 2020. Recommendations on practice to support FAIR data principles. FAIRsFAIR D3.4. <https://doi.org/10.5281/zenodo.3780423>

⁶⁴ See e.g. Garcia L., Batut B., Burke ML., Kuzak M., Psomopoulos F., Arcila R., et al., 2020. Ten simple rules for making training materials FAIR. PLoS Comput Biol 16(5): e1007854. <https://doi.org/10.1371/journal.pcbi.1007854>

⁶⁵ Newbold, E., Kayumbi Kabeya, G., Matthews, B., Davidson, J., Herterich, P., Whyte, A., Molloy, L., 2020. Initial Core Competence Centre Structures. FAIRsFAIR D6.2. <https://doi.org/10.5281/zenodo.3732889>

⁶⁶ <https://www.curationexchange.org/>

Recommended actions:

- Develop an understanding of costs of your service, such as costs for curating a dataset or implementing a new feature or standard supporting FAIRness.
- Make any costs associated with deposit clearly visible so that they can be easily found and included in grant applications.

What FAIRSFair will provide:

- Support to ensure that the costs of making and keeping data FAIR are better understood across the entire lifecycle and various stakeholders.

2.10 Benefit from engaging with the European Group of FAIR Champions network

FAIRSFair is proud to have initiated the European Group of FAIR Champions.⁶⁷ The group works as an ambassador of FAIR by sharing FAIR implementation stories, enhancing synergies, contributing to training activities and webinars, and doing an effective cross fertilisation with other communities, towards a broader engagement on FAIR. The FAIR Champions are scientific experts and “doers” in the field of FAIR data, carefully selected based on their individual merits and knowledge. Several of them are repository experts.

Recommended actions:

- Get in touch with a FAIR Champion to learn from them.

What FAIRSFair will provide:

- A forum for communication between the FAIR Champions and the wider community via the FAIR Competence Centre.

⁶⁷ <https://www.fairsfair.eu/advisory-board/egfc>