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

M3.5 Draft 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), Patricia Herterich (DCC), Angus Whyte (DCC), Josefine Nordling (CSC), René van Horik (DANS)
Due Date	30.06.2020
Date	08.07.2020
Version	0.5
DOI	https://doi.org/10.5281/zenodo.3935755
Link to working version	https://docs.google.com/document/d/1VXgrZZi23KrahihQ5Fp_Ym4IJ79FNKCWoridTzXlHOc/edit?usp=sharing

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

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, Herve L'Hours, Anusuriya Devaraju, Claudia Behnke, Hylke Koers, Elizabeth Newbold	Invited internal feedback and comments processed. Document finalised for public comment.

Disclaimer

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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Abbreviations and Acronyms

API	Application Programming Interface
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
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

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Executive Summary

FAIRsFAIR 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.

FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Table of contents

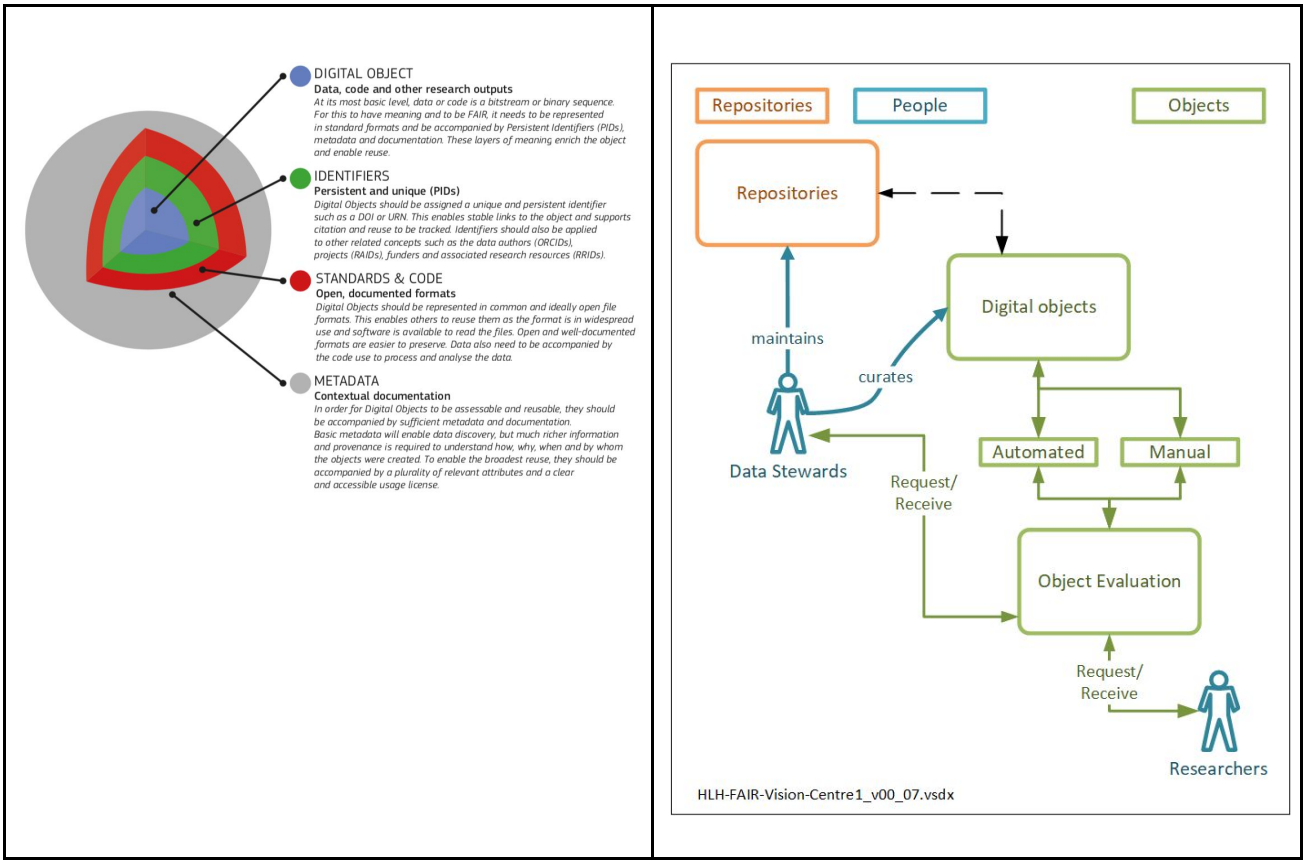
FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

1. Repositories and FAIR data: a shared journey

Digital repositories play an important role in making research data FAIR 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" report[1] and how FAIR digital objects fit into the wider repository ecosystem.[2]

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



[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. FAIRsFAIR. https://doi.org/10.5281/zenodo.3734273.

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

1.1 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 Approval[3] and ICSU-World Data Systems[4]), Extended (nestor/DIN31664[5]) and Formal (ISO16363[6]). In FAIRsFAIR, 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+FAIR[7] and is a work in progress.

Dillo and De Leeuw[8] and Donaldson et al.[9] examined the benefits of acquiring certification from the point of view of repositories that achieved the Data Seal of Approval[10], one of the two certification schemes from which the current CoreTrustSeal[11] (CTS) certification evolved. There are a number of benefits that can result from achieving certified status. Donaldson et al. report that the repository 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 Commission[12] 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.

---

[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] All versions of these emerging requirements can be accessed from https://doi.org/10.5281/zenodo.3734896

[8] 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

[9] 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

[10] https://datasealofapproval.org

[11] https://www.coretrustseal.org/

[12] 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)

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

1.2 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. Although the focus is on research data, repositories with other kinds of data, for instance Public Sector data, are encouraged to aim for enabling FAIR data and potentially to meet the requirements of a Trustworthy Digital Repository (TDR) by acquiring the CoreTrustSeal certification. After all, 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 longer-term preservation 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 over time. We encourage them to share any lessons they have learned with FAIRsFAIR 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 FAIRsFAIR project[13], 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 FAIRsFAIR 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 FAIRsFAIR will provide. This document will be shared for wide consultation and, based

[13] 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/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

on the feedback received, we will focus our efforts over the remainder of the project to develop support that will be most useful and has the greatest potential for realising change.

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 a registry service such as re3data offered by DataCite[14], which holds information on nearly 2500 research data repositories from across the globe. Thanks to the many filters offered, users can search for repositories that meet their specific needs. Through the FAIRsFAIR project, DataCite is developing a filter in their Repository Finder tool[15] to enable searching for repositories holding FAIR data content. The Repository Finder tool has been extended to query the re3data registry for repositories relevant to FAIRsFAIR 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. Your 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).

Recommended actions:

- Register your repository with re3data and other repository registries.
- Refine and/or develop clear policies relating to the repository service.
- Make policies visible to both humans and machines so that they can be put into action.

What FAIRsFAIR will provide:

- Guidance on ways to improve the findability of your repository's FAIR content - both for users within your specific domain and from other disciplines
- 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.

[14] https://datacite.org/

[15] https://repositoryfinder.datacite.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

2.2 Improve the interoperability potential of your repository content by becoming 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 features[16] 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 Point (FDP) is a repository for providing metadata access in a FAIR-enabling manner. FDPs have been implemented by several services, offering good examples of well-structured metadata mappings. These include B2SHARE[17] and EOSC-Pillar.[18] The software[19] 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.[20]

FAIRsFAIR cooperates with six 'developer repositories' and six 'tester repositories'.[21] The first step is to build a prototype with DCAT 2[22] 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 reference implementation of the repository interoperability features is planned to be ready in February 2021 and the second implementation one year later.

[16] 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
[17] 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
[18] https://www.eosc-pillar.eu/establishing-fair-data-services
[19] https://github.com/FAIRDataTeam/FAIRDataPoint
[20] FAIR Data Point specification: https://github.com/FAIRDataTeam/FAIRDataPoint-Spec
[21] The developer and tester repositories are listed here: https://www.fairsfair.eu/application-results-open-call-data-repositories
[22] https://www.w3.org/TR/vocab-dcat-2/

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

Recommended actions:

- Support the description of research outputs using agreed vocabularies, terminologies and metadata standards.
- 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 current supports these.
- Review the FAIR Data Point specification and its potential relevance for your repository.
- Consider setting up an FAIR Data Point in order to establish a common method for metadata provisioning and accessing.

What FAIRsFAIR will provide:

- Guidance for repositories on how to assess their readiness to support FAIR data features
- A reference implementation of the repository interoperability layer in a open source format
- A sandbox environment for testing the reference implementation
- Guidance on how to set up your own FAIR Data Point

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

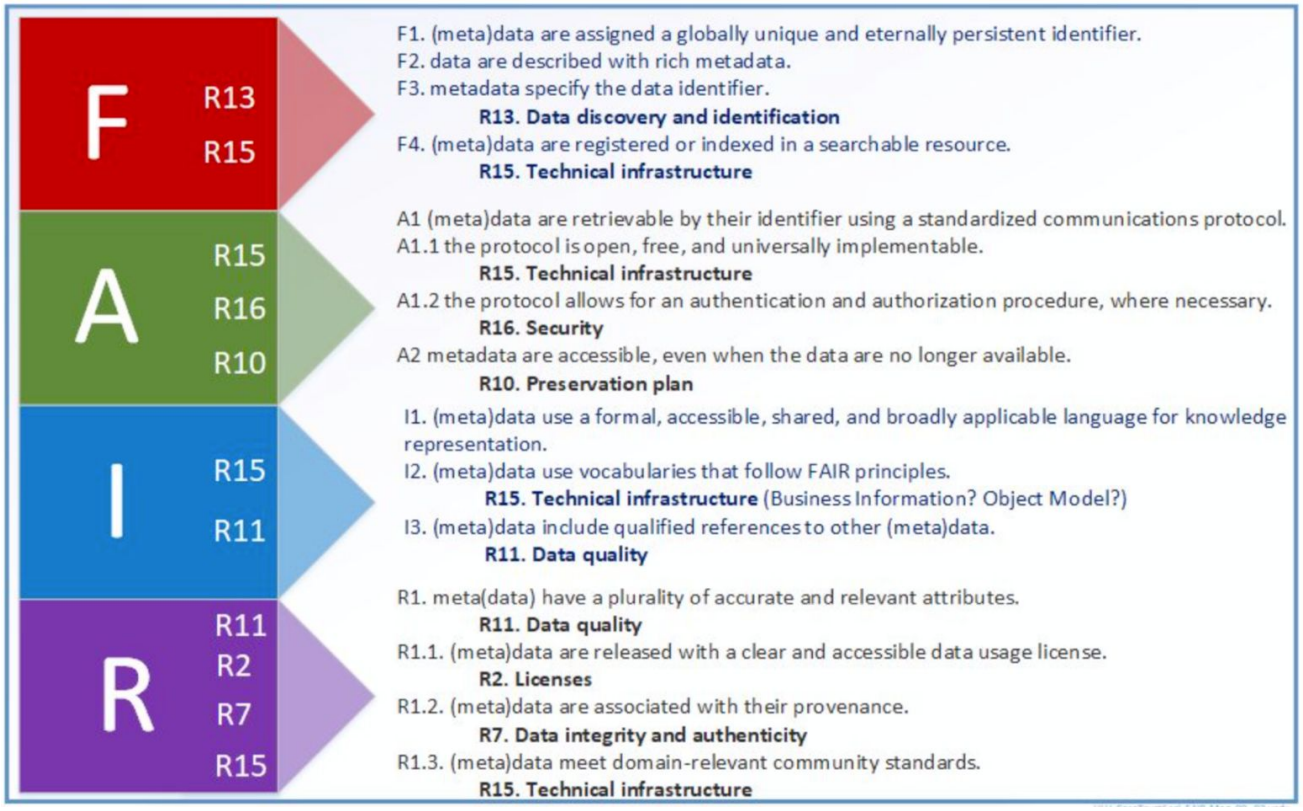
Starting from the notion that *data* should be FAIR, the question arises to what extent and how data *services* such as repositories can support the provision of FAIR data. For data services in general, Bangert et al. make recommendations that are based on a series of community workshops.[23] For repositories 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 principles[24], where “R” notes a CTS requirement.

[23] Bangert, D., Hermans, E., van Horik, R., de Jong, M., Koers, H., Mokrane, M., 2019. Recommendations for Services in a FAIR data ecosystem. https://doi.org/10.5281/zenodo.3585742

[24] 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

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. 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.

FAIRsFAIR 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. This will lead to recommendations for future iterations of the CTS requirements. An updated version of "CoreTrustSeal plus FAIR overview" is available.[25] Work on CoreTrustSeal+FAIR and on FAIR assessment will be used to inform recommendations on object and repository metadata.

All interested data repositories are encouraged to consider certification against the CoreTrustSeal's requirements for Trustworthy Digital Repositories, see the website[26] and the short video.[27] Even if

[25] 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
[26] https://www.coretrustseal.org/
[27] Walker, E., Ditsen de Jong, Y., De Bruin, K. and Chatzipoufli, N.D., 2019. CoreTrustSeal - Awareness-subs. 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

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

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 CoreTrustSeal certification.[28]

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 metrics are available for comment on Zenodo.[29] 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 work on tool implementations of the metrics and address FAIR assessment scenarios.[30] The first planned use case focuses on raising awareness of researchers who can self-assess the FAIRness of their data manually before depositing them into a repository. Although the tool will be hosted by DANS it will be generic and can be used for self-assessment of data held in other repositories. A prototype version of the tool (named

[28] 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

[29] 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

[30] 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

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

FAIR-Aware) is open for consultation until July 12 2020[31] and the following version of the self-assessment tool will be published in August 2020. The second planned use case focuses on the automated assessment of data objects deposited and published in selected data repositories.[32] For this use case, PANGAEA develops a web service (named F-UJI) 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 license. The results of the automated assessment will be shared through a planned journal article in August 2020.

Recommended actions:

- Become familiar with the FAIR 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.

What FAIRsFAIR will provide:

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

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.[33] 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.[34] Standardised descriptions for data sharing exceptions in metadata schemas could help pave the

[31] https://fairaware.dans.knaw.nl/
[32] https://github.com/pangaea-data-publisher/fuji
[33] 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
[34] As recommended in Davidson, J., Grootveld, M., Whyte, A., Herterich, P., Engelhardt, C., Proudman, V., Stoy, L., 2020. Policy Enhancement Recommendations. FAIRsFAIR D3.3. https://doi.org/10.5281/zenodo.3686901

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

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 which describes roles and responsibilities.[35] 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".[36]

Recommended actions:

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

What FAIRsFAIR will provide:

- 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)".[37] 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

[35] As outlined on page 11 of the FAIRsFAIR 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. FAIRsFAIR D2.3. https://doi.org/10.5281/zenodo.3631528

[36] https://support.datacite.org/docs/tombstone-pages

[37] 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

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

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, and organisations. 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)[38] - a community-led project to develop an open, sustainable, usable, and unique identifier for every research organisation in the world.

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.[39]

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 project[40] and OpenAIRE.[41]

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.

[38] https://ror.org/about/

[39] The first version of the guides to choosing Persistent Identifiers can be found at: https://zenodo.org/record/3862656

[40] https://www.project-freya.eu/en/pid-graph/the-pid-graph

[41] 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

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

2.7 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 can inform the repository about anticipated storage and metadata requirements, and be a trigger for the repository to timely inform the researcher, for instance about preferred file formats and available licences. 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 group[42], 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 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.[30]
- Get involved with the creation of DMPs and consider how your ingest and data management processes could benefit from information contained in DMPs
- 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.[43]

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 longer-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 might even get involved in it, for the benefit of

[42] Miksa, T., Walk, P., Neish, P., 2019. RDA DMP Common Standard for Machine-actionable Data Management Plans. https://doi.org/10.15497/rda00039
[43] https://rd-alliance.org/groups/exposing-data-management-plans-wg

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 grant agreement 831558

their own staff (as participants) and in the training (as tutors).[44] FAIRsFAIR is building on the successful CODATA/Research Data Alliance schools model[45], 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 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.[46]

Recommended actions:

- Explore FAIRsFAIR training materials.
- Consider how to make your training and learning material FAIR.[47]
- 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.[48]
- 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 longer-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 are needed 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. Developing a clear

[44] https://www.fairsfair.eu/events/training
[45] https://codata-rda-datascienceschools.github.io/DataSteward/
[46] 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
[47] 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
[48] 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

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

catalogue of costs for services will help to ensure that researchers - and more likely RPOs - can factor these into grant proposals as eligible costs.

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 t 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.[49] 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.
- Apply for the next call for FAIR Champions[50] - deadline 31 August 2020.

What FAIRsFAIR will provide:

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

[49] https://www.fairsfair.eu/advisory-board/egfc
[50] https://www.fairsfair.eu/form/open-call-european-fair-champions

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. Next steps

This brief description outlines some of the actions that can help repositories on their journey towards better alignment with the FAIR data principles and the support that FAIRSFAIR could provide. Please help us to shape the work we will do over the remainder of the project to develop and provide support for repositories of all types. We eagerly look for your feedback on:

- the recommended actions - do these reflect your repository mission and aims? Have we missed something?
- the support that FAIRsFAIR could provide - are some of the suggested support areas more helpful for your repository than others? Is there anything missing?

Please share your feedback with us by August 21, 2020 by adding comments to the working version using 'Suggesting' mode. The working version is available at: https://docs.google.com/document/d/1VXgrZZi23KrahihQ5Fp_Ym4IJ79FNKCWoridTzXlHOc/edit?usp=sharing

From the feedback we receive, FAIRsFAIR will prioritise its activities to develop guidance and provide support. The prioritised plan will be shared in the final version of the support programme in October 2020 (D3.5 Transition Support Programme for Repositories).

If you are interested in collaborating with FAIRsFAIR to review and/or use our support materials as they are developed, please contact us at https://fairsfair.eu/form/contact-us.

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