FAIRSFAIR
Fostering Fair Data Practices in Europe

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

M4.5 EVALUATION OF PROCEDURES AND PROCESSES OF CERTIFICATION MECHANISM PROVIDED

Work Package	WP4, FAIR Certification
Lead Author (Org)	Ilona von Stein, Frans Huigen (DANS), Hervé L'Hours (UKDA)
Contributing Author(s) (Org)	Olivier Rouchon (CINES), Jerry de Vries (DANS), Patricia Herterich (DCC)
Due Date	31.03.2020
Date	31.03.2020
Version	1.0

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

In this milestone, the issue of digital repository trustworthiness is addressed. A great variety of national and international standards, as well as guidelines for what constitutes a Trustworthy Digital Repository, exists. To formalize the interpretation and evaluation of FAIR principles and on the question how repositories can demonstrate their FAIRness, an assessment process has to be designed. In order to do so, this milestone evaluates existing certification mechanisms and procedures. Our recommendation is to consider the FAIR implications for CoreTrustSeal requirements by FAIR-aligning the Extended Requirements for the core certification of repositories.

Versioning and contribution history

Version	Date	Authors	Notes
0.1	10.03.2020	Ilona von Stein	Initial milestone content outline ready for consultation within WP4 biweekly meeting.
0.2	20.03.2020	Ilona von Stein, Frans Huigen, Hervé L'Hours	Process references gathered and reviewed, and first draft texts.
0.3	24.03.2020	Ilona von Stein, Frans Huigen, Hervé L'Hours, Olivier Rouchon, Jerry de Vries, Patricia Herterich	Initial draft ready for internal review within WP4.
0.4	27.03.2020	Ilona von Stein, Frans Huigen	Comments incorporated after WP4 internal review.
1.0	31.03.2020	Ilona von Stein, Frans Huigen, Hervé L'Hours	Processed the final comments, inserted diagram, final check. Submitted to PCO.

FAIRSFAIR
Fostering Fair Data Practices in Europe

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

FAIR	Findable, Accessible, Interoperable, Reusable
EOSC	European Open Science Cloud
TDR	Trustworthy Digital Repository
WP	Work Package
ISO	International Organization for Standardization

FAIRSFAIR
Fostering Fair Data Practices in Europe

Table of contents

1. Introduction

The aim of this milestone is to provide an evaluation of processes and procedures for FAIR-aligned repository evaluation and assessment in the European Open Science Cloud (EOSC). The FAIRsFAIR project is developing FAIR-aligned extended requirements for the core certification of repositories, and it is essential that this is supported by an evaluation process. The project considers how repositories can enable the FAIR data principles and how FAIR data characteristics affect core certification of repositories; the FAIR Principles on their own do not provide for an evaluation and assessment process for FAIR objects in FAIR-enabling data repositories. This milestone on processes and procedures for FAIR-aligned repository evaluation, is input to the initial version of a repository certification mechanism, including a FAIR-oriented elaboration of core Trustworthy Digital Repository (TDR) requirements, an initial draft of which will be released by the FAIRsFAIR project as deliverable 4.2 (due date May 2020).

2. Context

2.1 FAIRsFAIR Work package 4

This milestone is created in the context of Work Package (WP) 4 of the FAIRsFAIR project. This WP is focussed on the evaluation and certification of FAIR objects and FAIR-enabling repositories. It takes an iterative approach to testing and revision that will consider how FAIR object maturity can be aligned with existing core repository requirements, and where additional object-focussed criteria might be desirable.

2.2 Task 4.3 Support for FAIR-aligned certification

An evaluation of the processes and procedures of a repository certification mechanism is incorporated in FAIRsFAIR task 4.3 "Support for FAIR certification". This task provides support and capacity building towards FAIR-aligned certified repositories. Together with task 4.2, through a call for repository involvement,[1] it identifies and expands a European network of trustworthy repositories enabling FAIR data. The efforts in task 4.2 resulted in ten repositories being selected for in-depth evaluation and certification support. Tasks 4.3 and 3.4 are developing a transition support programme with a far wider range of repositories. Our contacts in these repositories share their knowledge and experience of the FAIR Principles as they relate to enabling FAIRness. This invaluable community feedback on repository practices, together with our FAIR-aligned certification support, allows for an iterative approach that we will continue throughout the development of a FAIR-aligned repository certification mechanism.

2.3 Interactions and dependencies

This milestone focuses on the evaluation of processes for repository certification in the EOSC. However, in line with the Turning FAIR into Reality report (European Commission Expert Group on FAIR data, 2018) FAIRsFAIR recognizes that in order for data and other research output to be FAIR a broader ecosystem of shared concepts, technologies, services, skills and culture is required. For a high-level vision of the FAIR ecosystem and its components interactions see (L'Hours and von Stein, 2019).

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

Within the FAIRsFAIR project, FAIR evaluation and assessment is addressed at different levels by different work packages, e.g. by WP2 at infrastructure level for services and software. There is a clear overlap with FAIRsFAIR task 2.4 that has released an initial report on the topic of assessing the FAIRness of services (milestone 2.7) and that will propose FAIR recommendations for software as well.

This milestone also has overlaps and dependencies with other EOSC-related initiatives. The EOSC-FAIR Working Group[2] will develop recommendations for implementing FAIR in practice based on the outcomes of selected projects, including FAIRsFAIR. FAIRsFAIR, OpenAIRE, EOSC-hub, FREYA and RDA Europe jointly organized workshops with different stakeholders to identify recommendations for services enabling FAIR data (Bangert et al., 2019). We will align our work with other criteria that might be defined for involvement in the EOSC, the five ESFRI clusters, the thematic and regional projects ('INFRAEOSC 5b') as well as on the five established EOSC Executive Board Working Groups. For WP4 there is a direct connection to the work in the EOSC Working Group on 'FAIR', especially to its sub-teams 'FAIR metrics' and 'FAIR Service Certification'. Through the FAIRsFAIR Synchronization Force, a dialogue among the various projects and actors in the EOSC ecosystem has already been established and will be continued to maximize coordination.

3. Process methodology for FAIR-aligned Extended Requirements for the core certification of repositories

Within FAIRsFAIR we take an approach to build on existing repository certification mechanisms and not reinvent the wheel. Therefore, we are neither proposing a new certification body nor a new procedure for repository certification.

To move beyond informal discussions on the interpretation and evaluation of the FAIR Principles and on how repositories can demonstrate that they enable FAIR we need to design and apply an assessment process. This work falls into two strands, first the internal project approach to self-assessment and peer review for supported repositories and second, recommendations for future operational systems which integrate trust and FAIR.

2 https://www.eoscsecretariat.eu/working-groups/fair-working-group

3.1 FAIRsFAIR Self-Assessment & Peer Review

With an extensive range of disparate evaluation approaches it's helpful to develop a structured typology of concepts and how they interact. This lets us design and evaluate evaluation standards and processes and compare them. We have developed a Generic Evaluation Reference Model (L'Hours and Bell, 2019). It provides a generic overview of key terms, concepts and functions around governing and implementing standard requirements and associated procedures.
In an assessment/evaluation, an entity (such as a digital object) is compared to a set of standard requirements, taking for example the FAIR principles. The FAIR principles are not sufficient on their own to support an evaluation process - in addition to the principles we require a governed process to undertake the review, a model like the GAERM (Generic Assessment and Evaluation Reference Model[3]) can provide a useful baseline for designing such an approach. In defining the enabling of FAIR data within a repository we must identify what can be assessed globally (across all data and metadata) and what clarifications of the principles are required. We must also identify principles that depend on some local context (e.g. domain-specific standards). Some standard description of what is sufficient to define context/clarification would be ideal. Context-specific extensions can then be created.

The RDA FAIR Data Maturity Working Group is extending the FAIR acronym into 15 principles, for which indicators, metrics, and tests[4]. An evaluation process in general compares the entity (object, repository, service etc) to the standard requirements; in this case in comparison to the FAIR indicators. Different evaluation processes can exist e.g. one process for global FAIR and another to cover data in the life science context.

Evaluation processes can use different assessment methods. The RDA FAIR Data Maturity Group qualifies these indicators as essential, important or useful. This might be considered more part of the standard requirements definition than as opposed to an assessment of maturity. Furthermore, scaling A to F (e.g. school grades), or a 1 to 10 rating provide some granularity to the assessment method, but don't in themselves define the outcome.

The outcome of an evaluation process could consider a D or a 5 as either a 'pass' or a 'fail'. The FAIR score of an ebook may lead to a different outcome from the same score for an object

3 https://zenodo.org/record/3733280
4 https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

containing sensitive personal data. Two evaluation processes which mark a number of answers on a scale from 1-10 might weigh the questions differently and result in different final scores.

Evaluation processes may have different outcomes. A pass in FAIR Training on sensitive data use may allow a researcher to use personal data forever, or require a refresher every 5 years, or last until personal data handling legislation changes. A trustworthy digital repository receives a CoreTrustSeal certification for three years after self-assessment and peer-reviewed evaluation of the actions and processes. But a digital objects' FAIRness is in doubt as soon as it has changed, for example because the DOI-reference has changed, or because the newer version fails a checksum.

3.2 Review of repository certification mechanisms

Regarding formal certifications, the International Organization for Standardization (ISO) publishes confirmed and refereed standards. Within the framework of information processing and trustworthiness of data, ISO 14721 exists. This ISO-standard is more commonly known as the Open Archive Information System (OAIS) model,[5] originally developed by The Consultative Committee for Space Data Systems (CCSDS). The Trustworthy Repositories Audit & Certification (in short: TRAC)[6] criteria are based on this OAIS-model. Further, ISO 14721/OAIS and TRAC lay the foundation for ISO 16363, which "(...) defines a recommended practice for assessing the trustworthiness of digital repositories" (ISO, 2017).

The Consultative Committee for Space Data Systems (CCSDS) Repository Audit and Certification Working Group also developed and submitted a second standard (the first one being ISO 16363), defining operational requirements for organizations intending to provide repository auditing and certification as specified in ISO16363. That second standard is published as ISO 16919: "Requirements for bodies providing audit and certification of candidate trustworthy digital repositories" (ISO, 2014).

These two formal standards are quite concrete examples of auditable processes that concern the harnessing of trustworthiness of and within repositories. There are many more ISO-standards being developed and published, of which the outputs of the ISO Technical Committee 171 are

5 ISO14721:2012 (CCSDS 650.0-P-1.1) https://www.iso.org/standard/57284.html
6 Trustworthy Repositories Audit & Certification: Criteria and Checklist: http://www.dcc.ac.uk/resources/repository

very much relevant, but out of scope for this milestone. Another ISO standard that is worth mentioning, but also out of scope for certification mechanisms, is ISO 17021. This standard, and underlying parts, is under patronage of the ISO Committee on Conformity Assessment. In contrast to the formal standards mentioned above, the ISO 17021 concerns certification bodies, rather than the organizations receiving certification. This ISO 17021 standard relates further to the way an audit is performed. Specific standards and procedures apply to an audit process, under the ISO 16919. This ISO is designed to be following a process striving for continuous improvement (ISO, 2014).

Next to the ISO-standards and related methods above, a wide variety of other national and international guidelines for digital data archiving do exist. For example, in Germany, the Network of Expertise in Long - Term Storage of Digital Resources awards the nestor Seal against their Criteria for Trustworthy Digital Archives (Kriterienkatalog vertrauenswürdige digitale Langzeitarchive)[7] (DIN31644). Five data repositories are currently evaluated against the DIN31644 standard. Another example is the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)[8] published by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE). The CoreTrustSeal,[9] launched in 2017, maintains 16 CoreTrustSeal TDR Requirements[10] and a certification process defined by a self-assessment and peer review workflow (see diagram 1 on the next page)[11] It offers core level certification for TDRs holding data for long-term preservation.

7 Zertifizierung, Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, https://doi.org/10.18452/1523
8 http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora
9 CoreTrustSeal website: https://www.coretrustseal.org
10 CoreTrustSeal Data Repository Requirements 2020-2022: https://doi.org/10.5281/zenodo.3638211
11 CoreTrustSeal Foundation Statutes and Rules of Procedures: https://doi.org/10.5281/zenodo.1142960

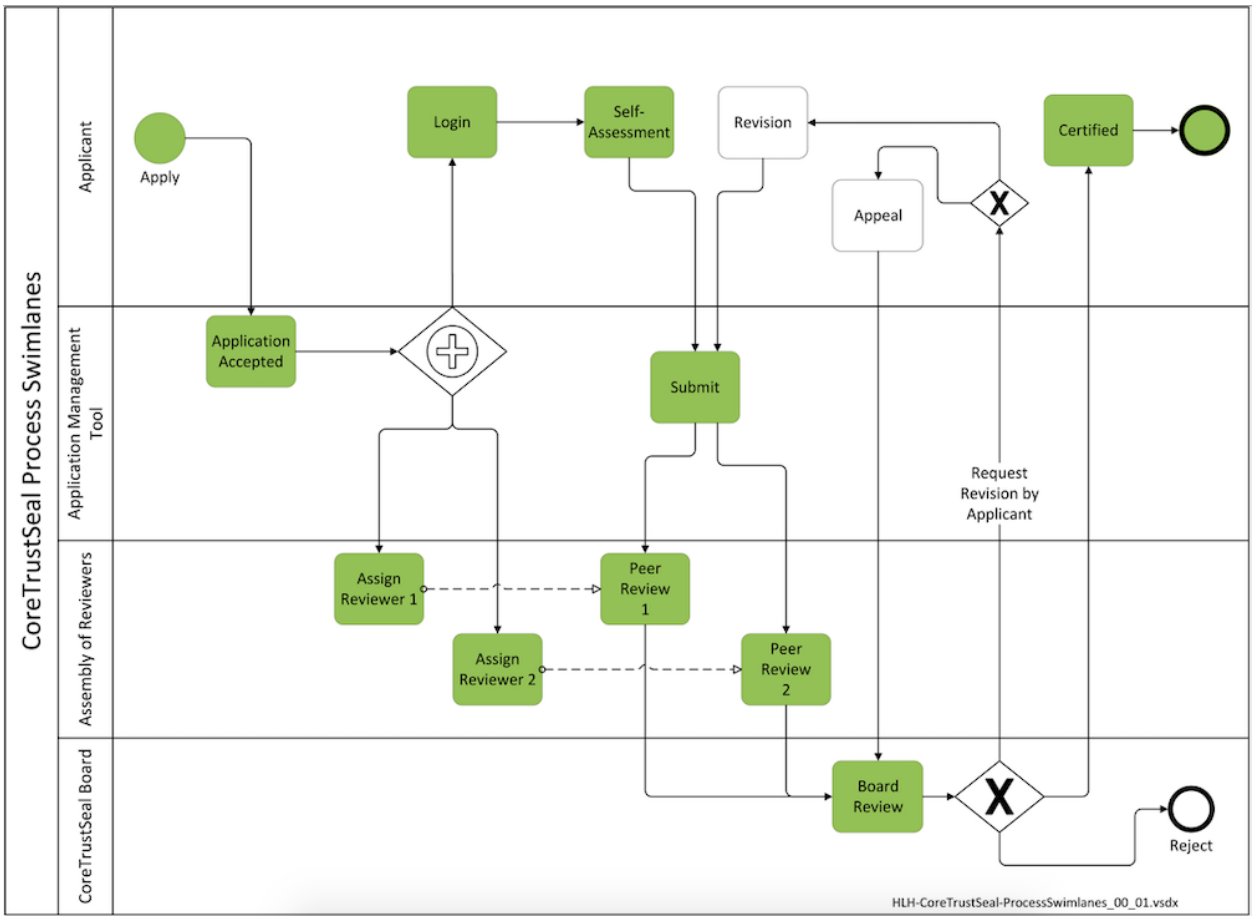

Diagram 1: CoreTrustSeal Process Swimlanes

3.3 FAIRsFAIR process proposal for CoreTrustSeal+FAIR

Ths brief overview of existing TDR standards makes it that there area range of possibilities available, published by various institutions, committees and working groups. However a standard needs a process too. The process for FAIRsFAIR's repository certification mechanism will build upon the CoreTrustSeal. Not only is it in line with the Turning FAIR into Reality report (European Commission Expert Group on FAIR data, 2018), but also CoreTrustSeal can be instrumental in helping repositories to adhere to the FAIR principles. The community-driven nature of CoreTrustSeal aligns strongly with the approach taken to develop the application of the FAIR Principles to operational work.

CoreTrustSeal is not only 'core' because it seeks to cover all the basic TDR requirements. It is 'core' because it tries to retain a level of structural simplicity and usability. In terms of effort, the

CoreTrustSeal certification is entry level and - like the FAIR principles - based on Research Data community needs. Formal certification like ISO is costly, bearing in mind the regulatory audits and dedicated certification processes. For example, the full ISO-documentation is placed behind a paywall. In contrast, CoreTrustSeal is completely openly available on their website, and the certification process is self-assessed and peer-reviewed. This fundamental difference - being open versus semi-open, is quintessential when working with FAIR data.

In FAIRsFAIR we follow an edited internal version of the CoreTrustSeal procedure internally (self-assessment and peer-review) for supported repositories, see also section 2.2. We will provide a repository support programme including workshops, webinars and an internal test-peer review of their repositories' CoreTrustSeal self-assessment, while at the same time we will draw on their expertise and knowledge of the FAIR Principles as they relate to repositories enabling FAIRness. The validation of this model exists in other EOSC contexts like CESSDA and SSHOC.[12] Our internal project approach lets us align with these and other actors that are together working on integrating trust and FAIR into the EOSC. This model will be iterated along with the standard requirements for CoreTrustSeal+FAIR. We will gather input from our experiences with the supported repositories, as well as from other proposed requirements for repositories and objects. Also we will iteratively evaluate our model against the Generic Assessment Evaluation Reference model as described in section 3.1 for consistency and gaps.

4. Conclusions

In order to address the issue of digital repository trustworthiness, there is a great variety of national and international standards and guidelines for what constitutes a Trustworthy Digital Repository. However, standards need processes too. To move beyond informal discussions on the interpretation and evaluation of the FAIR Principles and on how repositories can demonstrate that they enable FAIR we need to design and apply an assessment process. Comparing standards for TDR highlights both similarities and differences. FAIRsFAIR proposes CoreTrustSeal+FAIR as most valuable approach. Our recommendation is to consider the FAIR implications for CoreTrustSeal requirements by FAIR-aligning the Extended Requirements for the core certification of repositories. The first FAIRsFAIR proposal for CoreTrustSeal+FAIR requirements including process management and implementation will be released at the end of May 2020.

12 See e.g. CESSDA's overview of support approaches: https://doi.org/10.5281/zenodo.3621378

FAIRSFAIR
Fostering Fair Data Practices in Europe

Bibliography

Bangert, Daniel, Emilie Hermans, René van Horik, Maaike de Jong, Hylke Koers, and Mustapha Mokrane. 2019. 'Recommendations for Services in a FAIR Data Ecosystem'. Zenodo. https://doi.org/10.5281/zenodo.3585742 .

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

International Organization for Standardization (2014). ISO16919:2014, Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories. https://www.iso.org/standard/57950.html

International Organization for Standardization (2017). ISO16363:2012, Space and information transfer systems - Audit and certification of trustworthy digital repositories. https://www.iso.org/standard/56510.html

L'Hours, Hervé, and Darren Bell. 2019. 'Generic Evaluation Reference Model'. Zenodo. https://doi.org/10.5281/zenodo.3243154

L'Hours, Hervé, and Ilona von Stein. 2019. 'FAIR Ecosystem Components: Vision'. Zenodo. https://doi.org/10.5281/zenodo.3565428