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

D4.2 REPOSITORY CERTIFICATION MECHANISM: A RECOMMENDATION ON THE EXTENDED REQUIREMENTS AND PROCEDURES

Work Package	WP4 FAIR-Certification
Lead Author (Org)	Hervé L'Hours (UKDA)
Contributing Author(s) (Org)	Ilona von Stein, Jerry de Vries, Frans Huigen, Mustapha Mokrane, Linas Cepinskas (DANS), Anusuriya Devaraju, Robert Huber (UniHB), Joy Davidson, Patricia Herterich (DCC)
Due Date	31.05.2020
Date	31.05.2020
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.3835698

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

Aligning the CoreTrustSeal Requirements with an assessment of repositories' ability to enable FAIR data is an important part of delivering an EOSC. Trustworthy Digital Repositories which enable FAIR data are a dependency for many components of modern, open, distributed research. This paper sets the work within the wider context of data infrastructures, describes the co-dependencies between (meta) data objects and their repository environment, and presents the developing mapping between requirements and principles. The evolving capability/maturity approach is explained and the design of a governed assessment and certification process is defined. This work will iterate alongside the wide range of ongoing data infrastructure initiatives to support a range of stakeholders on their journey towards trustworthy repository services that enable FAIR data. Extensive engagement and feedback are planned to allow us to reach this goal.

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

Versioning and contribution history

Version	Date	Authors	Notes
0.1	2020-01-30	Hervé L'Hours (UKDA), Ilona von Stein, Jerry de Vries, Frans Huigen, Mustapha Mokrane, Linas Cepinskas (DANS), Anusuriya Devaraju, Robert Huber (UniHB), Joy Davidson, Patricia Herterich (DCC)	Draft for internal review.
0.2	2020-05-20	Hervé L'Hours (UKDA), Ilona von Stein, Frans Huigen (DANS)	Amendments in response to internal review.
1.0	2020-05-29	Hervé L'Hours (UKDA), Ilona von Stein, Jerry de Vries, Frans Huigen, Mustapha Mokrane, Linas Cepinskas (DANS), Anusuriya Devaraju, Robert Huber (UniHB), Joy Davidson, Patricia Herterich (DCC)	Content ready.

Disclaimer

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

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

Abbreviations and Acronyms

FAIR	Findable, Accessible, Interoperable, Reusable
TDR	Trusted Digital Repository
OAIS	Open Archival Information System
ISO	International Organization for Standardization
CMMI	Capability Maturity Model Integration
RDA	Research Data Alliance
WG	Working Group
EOSC	European Open Science Cloud
DDI	Data Documentation Initiative
PID	Persistent Identifier

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

This paper is Deliverable 4.2 of the FAIRsFAIR task 4.1 (Capability Maturity models towards FAIR Certification) within the FAIR Certification work package (WP4). The task will develop a practical and sustainable approach for repositories to self-assess their current capability levels and identify target levels for enabling FAIR data. This is the second step in aligning the characteristics of FAIR digital objects with the repositories that 'enable' FAIRness, through the CoreTrustSeal Trustworthy Data Repository Requirements[1] and the application of a capability/maturity evaluation approach: CoreTrustSeal+FAIR. The outcomes will be an overall improvement of repository practice and a pathway to certification.

The CoreTrustSeal is a community-driven effort to identify best practices, support improvement, and deliver better repository service outcomes to data users. The requirements and associated process are endorsed by the RDA[2] and have been explicitly recommended as the basis for certification of repositories by the Turning FAIR into Reality Report[3]. Certification offers recognition and demonstrates trustworthiness to data depositors, users and funders. However it is through the process of self-assessment and peer review that practices are shared and data infrastructures are improved. This FAIRsFAIR process follows that spirit of open inclusivity. The goal is to share and improve rather than exclude repositories or digital objects. Gaps in trustworthy repository practice or FAIR objects' status are opportunities for discussion and targeted improvement.

The individual goals of CoreTrustSeal, the FAIR principles, and the European Open Science Cloud (EOSC) align with an overall mission to maximise the quantity of FAIR data under trustworthy curation. Achieving this mission depends on actors working together to ensure that data are technically managed to ensure their protection and integrity, and preserved in a manner relevant to the types of objects and their user community. Ideally, digital objects also benefit from specialist preservation, e.g. by domain/subject experts such as disciplinary repositories.

The alignment of these requirements and principles must have operational value and be sustainable. CoreTrustSeal+FAIR will be iterated to support the evaluation of Trustworthy Data Repositories (TDR), including their ability to offer an environment that enables FAIR data and metadata for the long term.

A synopsis is provided, followed by the wider scope and context surrounding the work package, project, FAIR data and trustworthy data repositories. The methodology is described and the design principles of the proposed approach are outlined. Issues and dependencies are presented. The conclusion and next steps explain how the proposed approach will be opened to feedback and testing before a round of iterative updates.

[1] CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022
[2] https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs
[3] https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf Turning FAIR into Reality

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. Synopsis

The fifteen FAIR principles seek to set an expectation that digital objects (data and their associated metadata) become more findable, accessible, interoperable and re-usable. The RDA work to clarify indicators for the principles[4] has made it clear that a (digital) object cannot be made FAIR or evaluated for FAIRness in isolation from its context. Here, the relevant context is a data repository. The CoreTrustSeal is a community-driven foundation offering a certification process against sixteen core Trustworthy Data Repository (TDR) requirements. The FAIR and CoreTrustSeal approaches are complementary and well-aligned. The FAIR principles are neat, accessible statements about digital objects that also reflect the long-standing mission of repositories. Trustworthy Data Repository standards can enable FAIRness over time as they address changes to data assets and their users. A combination of FAIR and TDR offers an assurance that data will be preserved and continuously accessible for the long-term.

The ideal outcome of this work is a CoreTrustSeal process which certifies repositories as FAIR-enabling trustworthy data repositories. The clear alignment of CoreTrustSeal+FAIR has immediate benefits in addressing the relationship between data and users via repository data services. The challenge is to develop an aligned approach which offers both an assessment/certification mechanism and a useful tool that has operational value to good repository practices.

An assessment usually involves some evaluation/scoring method. In this case, we are using both the CoreTrustSeal compliance levels and a capability maturity model integration (CMMI) approach. CoreTrustSeal scores from 'not considered' to 'fully implemented' while CMMI scores from incomplete to optimising (see *CoreTrustSeal, Compliance, Capability & Maturity* below). We will evolve capability/maturity tiers for CoreTrustSeal+FAIR alongside the development of FAIR indicators.

Both defining and achieving FAIR are a journey. This aligns well with the CoreTrustSeal goal of providing clear expectations, with an assumption of improved repository practice over time. A transparent, supportive community of practice is best-placed to deliver a European Open Science Cloud (EOSC).

The users of data are implied but not directly addressed by FAIR. For example, the FAIR Reusable principle 1.3 "meet domain-relevant community standards" connects objects to users through repositories. TDR standards directly address the need to serve a defined community of users (see *Designated Community and Other Users* below).

In developing CoreTrustSeal+FAIR, a direct mapping of Requirements to the FAIR acronym is not sufficient. To align digital objects and the repository context, we must analyse the FAIR principles and the repository approach to data and metadata. This is also dependent on evidence for compliance

[4] https://www.rd-alliance.org/groups/fair-data-maturity-model-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

provided by repository process metadata and other business information (see *FAIR Object and FAIRenabling Environments* below). We must also consider the indicators of FAIRness which are still under development. These are necessary to identify metrics and to apply tests for FAIRness (See *CoreTrustSeal+FAIR: Draft Elaboration Model* below).

The agreement of indicators and the development of tests for FAIRness, including the degree of 'machine-actionable FAIRness' sit alongside the need to clarify FAIR concepts (e.g. the richness of metadata) and contexts (e.g. community standards). We monitor issues throughout our work (see *FAIR Principles: Baseline* under *Scope and Context* below)

Despite the clear alignment, there are two key challenges in designing CoreTrustSeal+FAIR:

1: Concepts that are implicit assumptions rather than explicit requirements in CoreTrustSeal.

2: Concepts in FAIR that have mappings to more than one part of the CoreTrustSeal.

Broadly speaking a repository may evaluate, curate and communicate for FAIRness at three points during the sequential repository phases (R8. Appraisal, R11. Data Quality and R14. Data ReUse). Together these create the environment for data discovery (R13) by the user. In CoreTrustSeal 'access' is assumed and implied through delivering a mission (R1) in line with licence conditions (R2). But all of the CoreTrustSeal Requirements remain critical to ensuring that organisations and objects are sustainable over time (preservation). See CoreTrustSeal: Requirements in Brief and CoreTrustSeal+FAIR below.

Clear, accountable assessment, evaluation and certification depend on a transparent and well-governed process. During the FAIRsFAIR support process, we will follow an amended version of the CoreTrustSeal Procedures. We will also be designing a recommended approach to CoreTrustSeal+FAIR in practice, for assessment/evaluation and eventual certification (see *A Governed Assessment and Evaluation Process* below).

As we iterate and collaborate with a wide variety of stakeholders the CoreTrustSeal+FAIR work will also integrate with the broader vision of an interoperable European Open Science Cloud (see *Wider EOSC Components* under *Scope & Contex*t).

The ideal outcome of this work is a CoreTrustSeal process, which certifies repositories as trustworthy data repositories that enable FAIR data. We have a design plan and an iteration process to test proposals for CoreTrustSeal+FAIR.

2. Scope & Context

The primary focus of this work is to align the CoreTrustSeal Requirements with FAIR to identify how repositories can enable FAIR data. Provision of a capability maturity approach is central to this work, but the application of capability and maturity levels will not be prescriptive at this stage. These will be developed iteratively through interaction with ten supported repositories and more extended

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

engagement, for example in the emerging European Network of Trustworthy Data Repositories enabling FAIR data.

The format of a single large deliverable is not best suited to addressing the complex content and varied stakeholder audiences. At this stage, the content is directed primarily at those designing and developing FAIR and EOSC related standards and infrastructures. These standards need to be streamlined and updated over time to provide clear direction to repositories and their key stakeholders: depositors, users and funders. The 'component documents' referred to will also evolve independently over the course of the project. This will lead to a final deliverable that proposes a standard, process and governance model that incorporates CoreTrustSeal+FAIR.

The outputs are not only directed at an operational repository audience, but also aimed at those designing interoperable infrastructures of people, processes and technologies. Making CoreTrustSeal+FAIR simpler and more usable for a wider range of stakeholders will form part of the FAIRsFAIR iteration process.

Within the FAIRsFAIR project work package 4 will: offer support for FAIR-enabling Repositories (T4.3), develop a network of FAIR-enabling Trusted Digital Repositories (T4.2), improve registries for FAIR-enabling repositories (T4.4) and undertake several FAIR Data assessment pilots (T4.5). These pilots and other work to formalise metrics and tests against the FAIR Principles will be used to evaluate how to or to integrate the FAIR 'scores' of repositories collections into FAIR-enabling repository assessment.

The FAIR Data Principles: Baseline

The detailed clarification of each principle and its application is beyond the immediate scope, though highly relevant to any final recommendations.

All current FAIR work can be traced back to the original 2014 Force 11 Principles and the subsequent Nature paper[5] which we use as our reference point. The numerous ongoing efforts around FAIR often question the meaning and intention of the original principles at different points in their work. We need to address these issues of FAIR interpretation without allowing them to delay our progress. We have annotated the Principles to develop a 'baseline' of potential issues (see Component Documents)[6] that impact the definition and evaluation of digital object FAIRness or the ability of repositories to enable their FAIRness. Each future iteration of CoreTrustSeal+FAIR will identify whether these baseline issues have been addressed.

Repository Interoperability

Interoperability between repositories and with other components of the EOSC is essential. This particularly applies to technical standards for repository interoperability. Full details of the FAIRsFAIR

[5] https://www.nature.com/articles/sdata201618 The FAIR Guiding Principles for scientific data management and stewardship
[6] https://doi.org/10.5281/zenodo.3728130 FAIR Principles: Baseline Comments

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

work in this area are presented in *D2.3 Set of FAIR data repositories features*[7]. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

Object Assessment

Among the many rapidly evolving areas of FAIR and EOSC is the ongoing development of indicators and tests for objects' compliance with the FAIR principles. Full details of the FAIRsFAIR work in this area, including interactions with the RDA FAIR Data Maturity Working Group, are available in the deliverable *4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories*[8]. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR. Draft recommendations from task 4.5 on what would be considered 'Core' requirements for FAIR data have been mapped and integrated.

Object & Repository Data

The CoreTrustSeal+FAIR alignment of repository practice with object assessment against FAIR supports FAIRsFAIR task 4.4 in identifying extensions to descriptive metadata about repositories. This sets the foundation for streamlining assessment and certification with improved organisational and data collection metadata.

Service Assessment

Repositories are part of a wider data service ecosystem. The FAIRsFAIR work in this area is available in the *Assessment Report on FAIRness of Services*[9]. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

Human Mediated and Machine-Actionable Assessment

The minimum expectations for machine-actionability will become more apparent as the FAIR Data Indicators are defined and tested, and as different aspects of the EOSC ecosystem mature. Future interactions of this work will take into account the expected balance of machine-actionability, including assessments across repositories, objects and services, partnerships and policies. This includes the evolving goals for semantic interoperability of repositories and machine-actionable policies.

Wider EOSC Components

The final recommendations from this work depend on repository interactions with the wider components of the EOSC Ecosystem. *FAIR Ecosystem Components: Vision*[10] is being iterated in response to external feedback and internal results. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

[7] https://doi.org/10.5281/zenodo.3631527 Set of FAIR data repositories features
[8] https://doi.org/10.5281/zenodo.3678715 Draft Recommendations on Requirements for FAIR Datasets in Certified Repositories
[9] https://doi.org/10.5281/zenodo.3688761 Assessment report on 'FAIRness of services'
[10] https://doi.org/10.5281/zenodo.3565427 FAIR Ecosystem Components: Vision

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

Policy Integration and Enhancement

The FAIRsFAIR Project provides a number of recommendations for policy enhancement[11] that will be considered as we develop repository assessment proposals. The key findings are structured in terms of the Turning FAIR into Reality report: define, implement, embed and sustain. Policy enhancements relevant to CoreTrustSeal+FAIR include "*Efforts are needed to raise general awareness about the FAIR principles and how to implement them in a practical sense*" (#1), "*Clearer definitions of data and expectations around sharing are needed. Definitions and expectations should be harmonised across stakeholders*" (#6-8), and "*Requirements for research data management (RDM) and data management plans (DMPs) should be harmonised across stakeholders*" (#14-18).

Assessment & Evaluation Modelling

The outcome of an assessment/evaluation of an object or other entity (such as a repository or service) is a defined status, e.g. Trustworthy, FAIR, Open. There are several existing and in development evaluation approaches for us to examine. A structured typology of relevant concepts facilitates the design, review and comparison of standards and processes[12]. Future iterations of the CoreTrustSeal+FAIR outcomes will be benchmarked against this model.

[11] https://doi.org/10.5281/zenodo.3686900 Policy Enhancement Recommendations

[12] https://doi.org/10.5281/zenodo.3243153 Generic Assessment & Evaluation Reference Model

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. FAIR Objects & FAIR Enabling Environments

Different scientific communities and their repositories work with different assumptions about what is a 'digital object' and different approaches to 'data' and 'metadata'. In *Turning FAIR Data into Reality*, the following overview object model is presented.

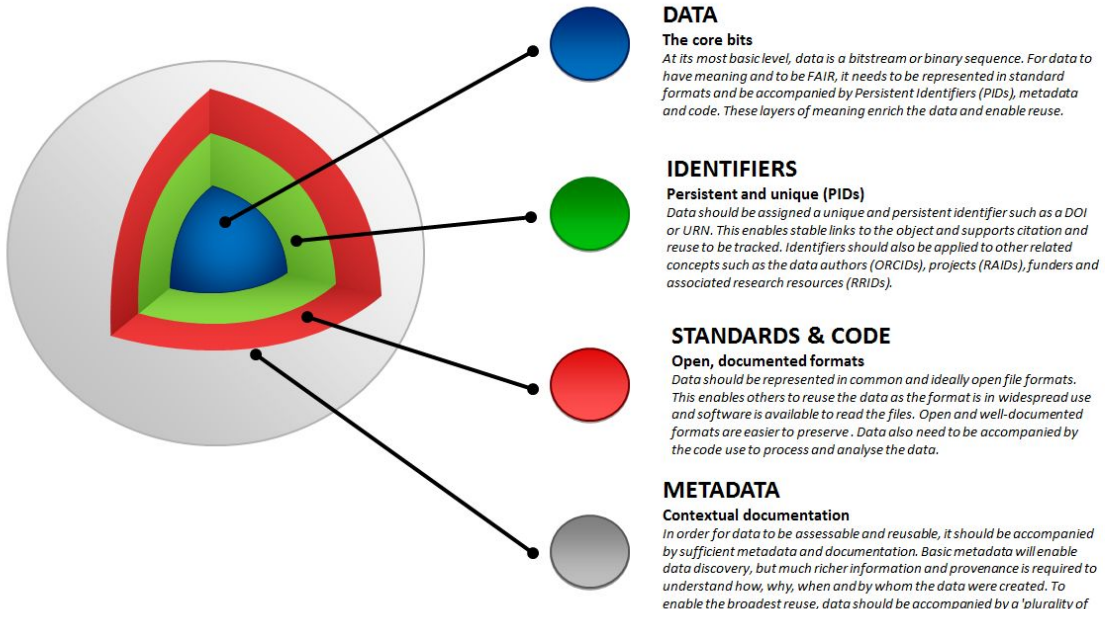


Diagram 1: ***Rec. 3****: A model for FAIR Data Objects*

This division between the data (as the original target for collection/creation) and its supporting metadata is not always clear and consistent in practice. For example, some standards support data and associated metadata contained within a single file (e.g. DDI[13], ABCD[14]). Repositories also create their own 'business information' which include policies, procedures and other documentation, and its own 'process metadata' (ranging from 'policy review/approval' to 'format risk updated'). Some of this repository 'process' metadata might be stored and managed with the object metadata (e.g. 'validation of a checksum' or 'file format migration completed'). All of these (meta) data types are important as either they enable FAIRness directly or they provide supporting evidence for enabling FAIRness.

The diagram below presents the potential overlaps between object data, object metadata, repository process metadata and other repository business information.

[13] https://ddialliance.org/Specification/ Data Documentation Initiative
[14] https://abcd.tdwg.org/ Access to Biological Collection Data

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

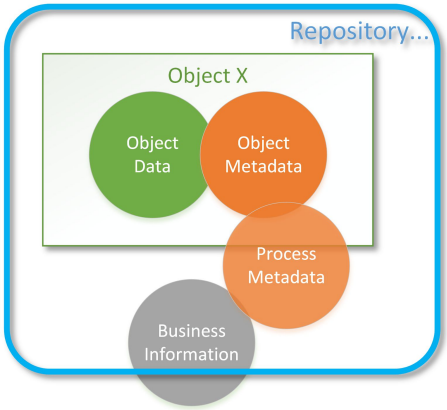


Diagram 2: ***Repository & Object Metadata***

In the development and implementation of CoreTrustSeal+FAIR, we must take into account repositories and their collections of heterogeneous digital objects. But we must also remain general enough for the approach to be applicable to a broad range of repositories. The diagram below demonstrates a mapping from objects to the FAIR principles that takes account of the repository context and some wider dependencies.

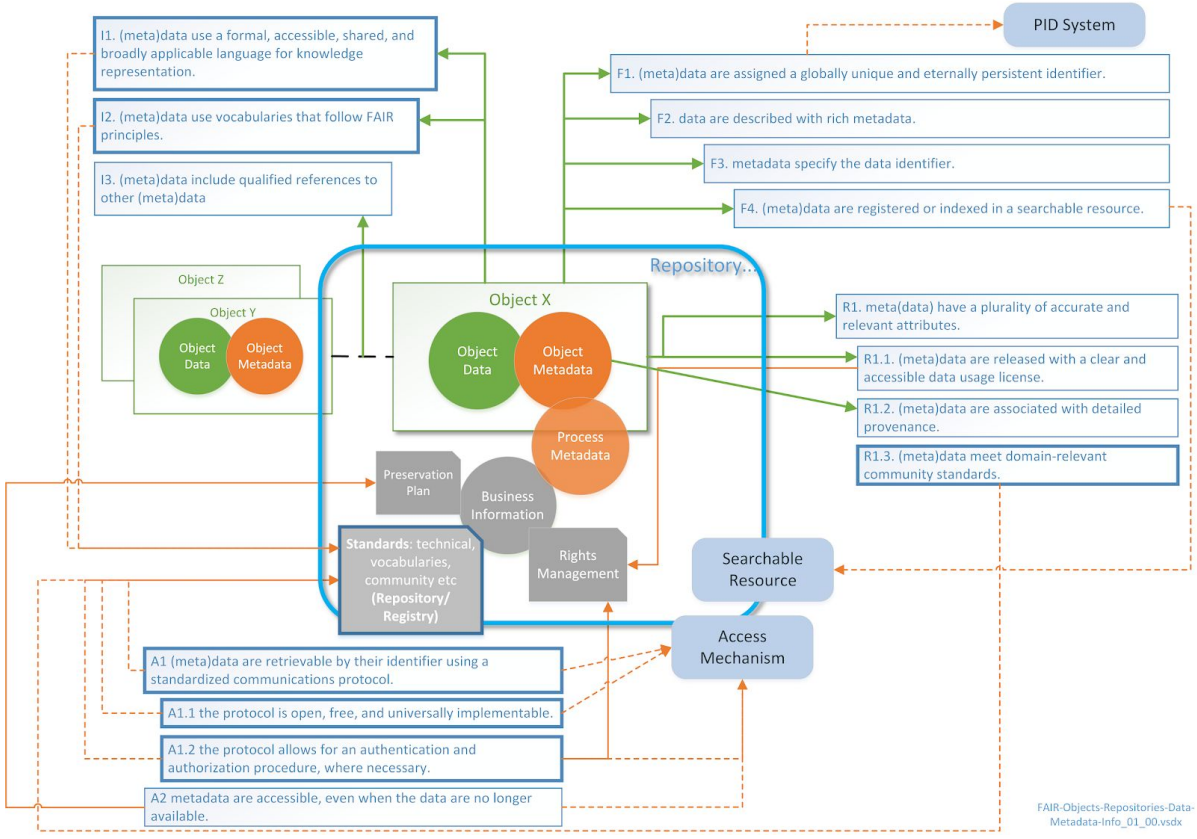


Diagram 3: ***FAIR Objects, Repositories, Dependencies*** *(FAIR Principles abbreviated)*

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

In the diagram above (full version see Appendix 1) the green arrows represent FAIR Principles that are most closely associated with object characteristics. However, delivering FAIRness remains dependent on the data curator. In this case, the repository is the data curator, but from a full lifecycle perspective FAIRness depends on data creators/researchers/depositors to provide FAIR data at source, and on data re(users) to follow FAIR principles. Orange arrows represent cases where compliance with FAIR Principles has dependencies, for example, on internal repository business information like rights management or preservation plans. Dotted orange arrows represent dependencies on functionality (PID systems, searchable resources, access mechanisms) or information (technical/community standards for data or metadata vocabularies) which might be outside direct repository control (e.g. held in a registry or provided as a third-party service).

Principles with a bold border indicate the (minimum number of) cases where there is a dependency on some wider clarification or contextualisation (e.g. “what is acceptable as ‘rich’ metadata?”, or “how must a vocabulary meet FAIR principles?”).

Defining the alignment between objects and their repository environments allows us to identify dependencies. It also helps us to identify which repository partners could provide supporting evidence for CoreTrustSeal+FAIR status.

3.1. CoreTrustSeal Requirements in Brief

The diagram below presents the CoreTrustSeal requirements. Context (R0) provides information to support the overall assessment. Organisation Infrastructure (R5), supports: internal expertise and governance, achieving the mission (R1), business continuity (R3), rights management (R2), confidentiality and ethical issues (R4) and access to appropriate external expertise (R6).

Digital Objects are preserved (R10) for ongoing access through selection and appraisal of deposits (R8), assurance of quality (R11) during curation and by enabling discovery (R13) and reuse (R14) through managed workflows (R12).

The integrity and authenticity (R7) of data and their storage (R9) are primarily addressed from the curator perspective in CoreTrustSeal, but they also depend on the Technical Infrastructure (R15) and Security (R16).

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

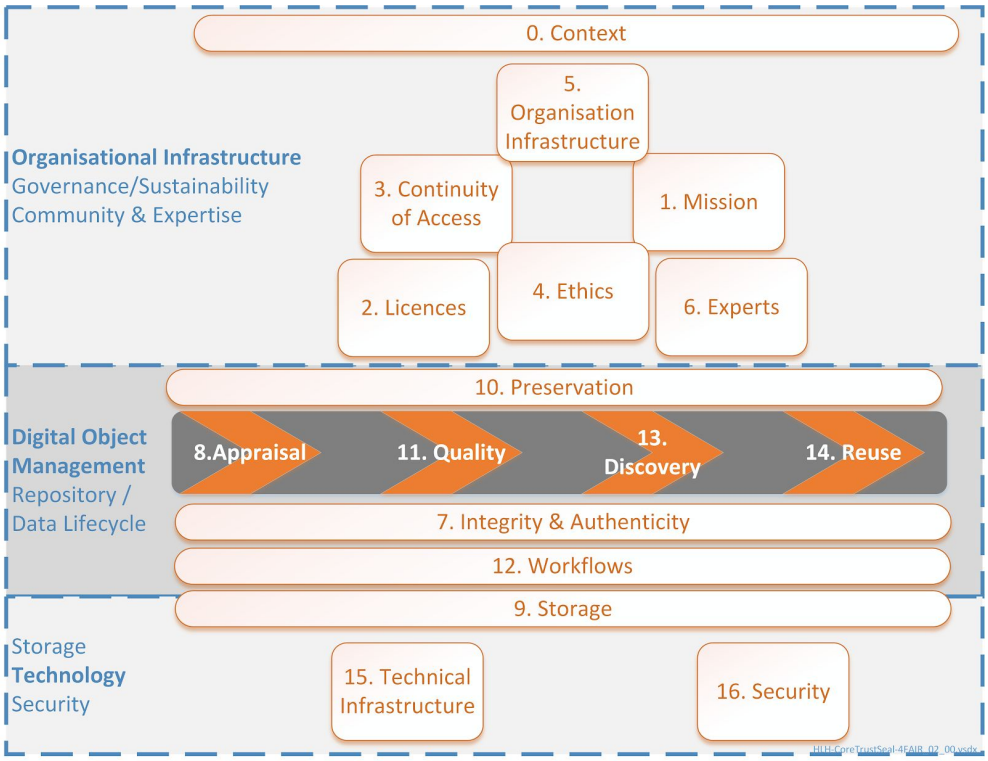


Diagram 4: ***CoreTrustSeal Requirements in Brief***

Broadly speaking a repository may evaluate/curate for FAIRness at three points

- R8. Appraisal
- R11. Data Quality
- R14. Data Reuse

Objects may be evaluated for FAIRness at *appraisal*. Curation to ensure *data quality* may apply missing elements of FAIRness. At the point of *data reuse,* the FAIRness of data should be assured, or any lack of FAIRness communicated to data users.

4. CoreTrustSeal+FAIR

In evaluating approaches to CoreTrustSeal+FAIR an initial mapping of the CoreTrustSeal Requirements to the FAIR acronym is a starting point.

- **Findable**: data discovery and identification (R13)
- **Accessible**: is explicit in Mission (R1) necessary for ReUse (R14)
- **Interoperability**: is a necessary condition to deliver reusability by the full range of stakeholders
- **Reusability** is implicit in the need to appraise (R8) quality assure (R11) for reuse (R14) by a clear community of users (R0)

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

The FAIRsFAIR CoreTrustSeal to FAIR alignment builds on previous work by CoreTrustSeal Board Members[15] at iPres and the OpenAIRE workshop on Services to support FAIR data[16]. The second iteration of the CoreTrustSeal to FAIR alignment mapping has been reviewed and responded to by the ten FAIRsFAIR Repositories supported within this FAIRsFAIR work package.

The *CoreTrustSeal+FAIR Overview*[17] presents high-level FAIR-related questions, asks for additional repository context, and maps the FAIR principles and the indicators being evolved by the FAIR Data Maturity Working Group[18] to the CoreTrustSeal Requirements.

There are areas where the requirements can be aligned directly with repository capability. In other cases, a single mapping is not possible as there are multiple areas of repository activity that contribute to FAIRness (e.g. Appraisal, Quality and Reuse). The overview is open to public comment, and the outcome of this feedback and review process will be a FAIR mapping integrated into a template of the CoreTrustSeal Extended Guidance.

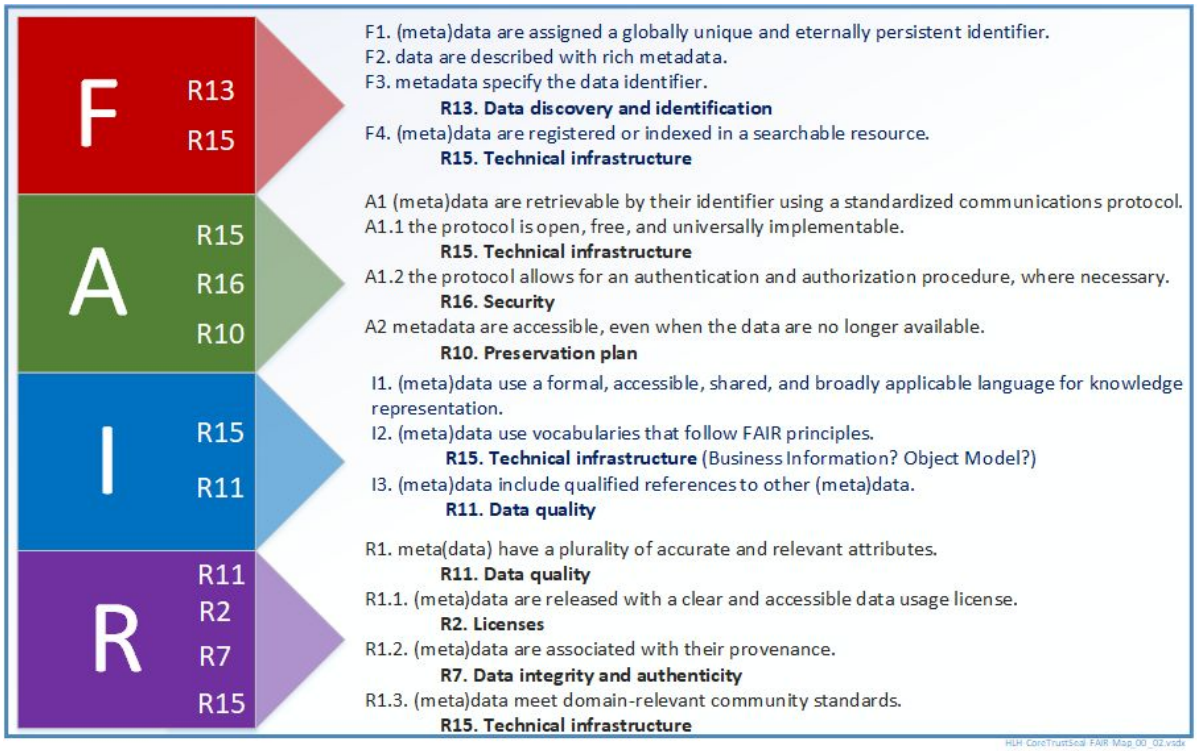


Diagram 5: ***FAIR to CoreTrustSeal***

A more detailed mapping from the Requirements to the Principles is provided in Appendix 2.

[15] https://ipres2019.org/static/pdf/iPres2019_paper_74.pdf Enabling Findable, Accessible, Interoperable, and Reusable (FAIR) Data

[16] https://www.slideshare.net/OpenAIRE_eu/how-core-trust-seal-enables-fair-data-natalie-harrower How CoreTrustSeal enables FAIR data

[17] https://doi.org/10.5281/zenodo.3734896 CoreTrustSeal plus FAIR Overview

[18] https://www.rd-alliance.org/groups/fair-data-maturity-model-wg FAIR Data Maturity Model 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

5. CoreTrustSeal+FAIR: Draft Elaboration Model

In setting up an approach for FAIR enabled repositories, we need to consider where we can elaborate on the existing CoreTrustSeal requirements and whether some additional features are required. The design methodology is to use the CoreTrustSeal Requirements as a baseline and to elaborate them in ways which demonstrate that a repository enables FAIRness.

The overall goal is to integrate the CoreTrustSeal requirements with repository approaches to enabling FAIR data. A capability/maturity approach will be used to support repository assessment and improvement. This will be aligned with parallel work to test the FAIRness of curated digital objects.

The repositories supported by FAIRsFAIR are the initial audience, but much more extensive feedback is sought as we iterate and test the approach. There are several logical mappings from FAIR into various parts of the requirements. However, we need to select the most intuitive and practical alignment, so repositories have clear locations to provide self-assessment statements and associated evidence for FAIR enabling.

The direct mapping of FAIR and CoreTrustSeal and the application of capability and maturity assessments (see below) have many challenges. The FAIR acronym contains 15 principles, each of which is under review to develop relevant indicators, metrics, and tests. The RDA FAIR Data Maturity Working Group are also classifying each indicator as one of: essential, important or useful.

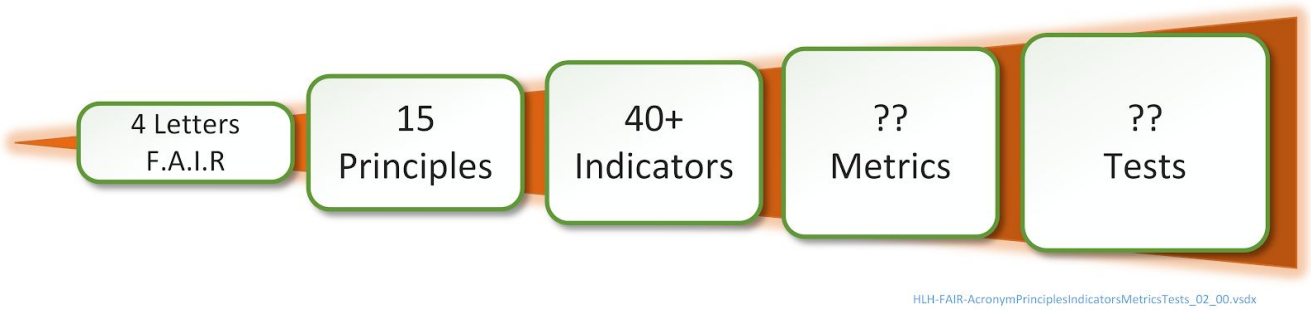


Diagram 6: ***FAIR acronym, principles, indicators, metrics & tests.***

Indicators clarify how the Principles apply in practice. Metrics define how that practice can be measured. A range of tests could be designed/coded to apply these metrics.

For example: "*R1.2: (Meta)data are associated with detailed provenance*." is supported by the indicator[19] "R1.2-01M Metadata includes provenance information according to community-specific standards. A metric would be the presence or absence of provenance information plus their alignment with an agreed community standard. The test could involve a search for provenance-related metadata

[19] https://doi.org/10.15497/RDA00045 FAIR Data Maturity Model: specification and guidelines

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

elements which comply with a provenance schema approved as community-specific and listed on FAIRsharing[20].

For the initial integration of CoreTrustSeal assessment with the FAIR principles, we have some open issues. These include the need for feedback from repositories about their perception of FAIR enabling and a more extensive set of contextual questions than those currently requested by CoreTrustSeal. There are also some FAIR concepts, including the use of standards and the provision of access functionality, which are implied by several CoreTrustSeal Requirements rather than being explicitly defined.

5.1. Supporting the Journey Towards Trust & FAIR

Both Trustworthy Data Repository status and FAIR data may be conceived as a journey. The application of a scoring mechanism such as capability/maturity may support repositories at lower levels of maturity in defining and achieving their goals. This approach can also be aligned with the work of the EOSC Secretariat working groups[21] including Rules of Participation[22] and FAIR which have resulted in *Interim recommendations for FAIR metrics and service certification to apply within EOSC*[23] and "Interim recommendations on certifying the services required to enable FAIR research outputs within EOSC"[24].

5.2. CoreTrustSeal, Compliance, Capability & Maturity

An assessment usually involves some evaluation/scoring method. In this case, we are using both the CoreTrustSeal compliance levels and a capability maturity model integration (CMMI) approach.

CoreTrustSeal Self-Assessment Compliance Levels

0 – Not applicable
1 – The repository has not considered this yet
2 – The repository has a theoretical concept
3 – The repository is in the implementation phase
4 – The guideline has been fully implemented in the repository

CMMI Levels

0: Incomplete	1: Initial	2: Managed	3: Defined	4: Quantitatively Managed	5: Optimizing

[20] FAIRsharing.org

[21] https://www.eoscsecretariat.eu/eosc-working-groups EOSC Working Groups
[22] https://www.eoscsecretariat.eu/draft-eosc-rules-participation-rop-feedback-survey Rules of Participation
[23] https://repository.eoscsecretariat.eu/index.php/s/C3a5WkpsFHL6GD3 Recommendations on FAIR Metrics
[24] https://repository.eoscsecretariat.eu/index.php/s/zCnHTcytBHaLjRp Recommendations on Certifying Services

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

Rather than pre-defining expectations against each aspect of CoreTrustSeal+FAIR, we will evolve our approach to these tiers over time. This will happen through interactions with the ten FAIRsFAIR supported repositories, the network of FAIRenabling Trustworthy Data Repositories and the global network FAIR and CoreTrustSeal stakeholders.

There are two reasons for taking an iterative and evidence-based approach. The first is that neither CoreTrustSeal nor FAIR are designed with CMMI in mind. They do not apply the practice areas which are mapped to assess practice capability and overall institutional maturity. The second is that the indicators to identify FAIRness, and the tests against those indicators are still under development. Outcomes of that activity will change the 'minimum' capability-maturity expectations and CoreTrustSeal+FAIR interactions more broadly.

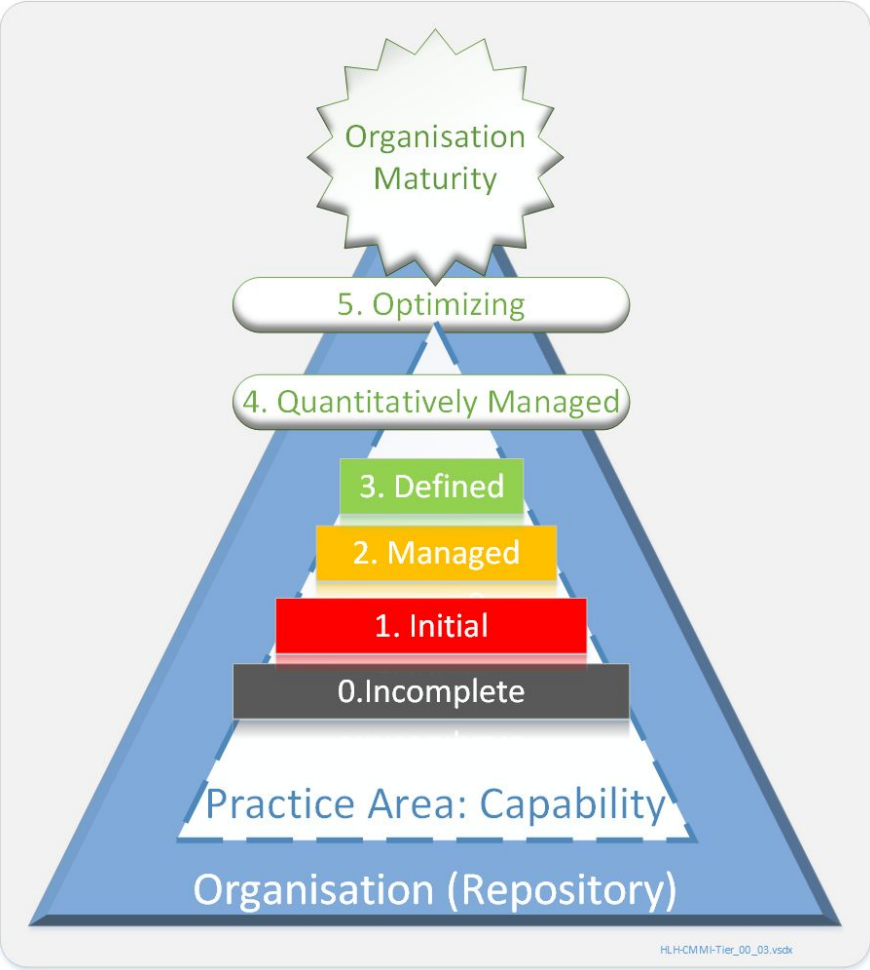


Diagram 7: ***Tiered Capability/Maturity***

For capability/maturity our working assumption is that capability levels of defined (3) can deliver FAIRness, though we will consider the validity of level 2 (managed). Maturity level 4 (quantitatively managed) may be a dependency for sustainable complex partnerships between data service providers.

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

CMMI is an operational tool and not a marketing device; achieving level 5: Optimising should be seen as desirable, but resource-intensive. It is valuable to support data services in defining where they need to focus resources on improvement. For further detail, see *Capability-Maturity Modeling and Landscape* in *Component Documents* below.

5.3. A Governed Assessment and Evaluation Process

5.3.1. Assessment Methods & Outcomes

It seems inevitable that there will be a debate on what constitutes a level 3 maturity (defined) vs level 5 (quantitatively managed) and on what outcome is required for a given set of circumstances (e.g. 3 for low value, low cost/easy to recreate data, 5 for high value or sensitive data). We expect community expectations to evolve. We also need to be sure the measurement/metric (e.g. CMMI scale) is appropriate to the object characteristics or repository features being analysed.

This assessment and evaluation must be applied through a governed and transparent review process.

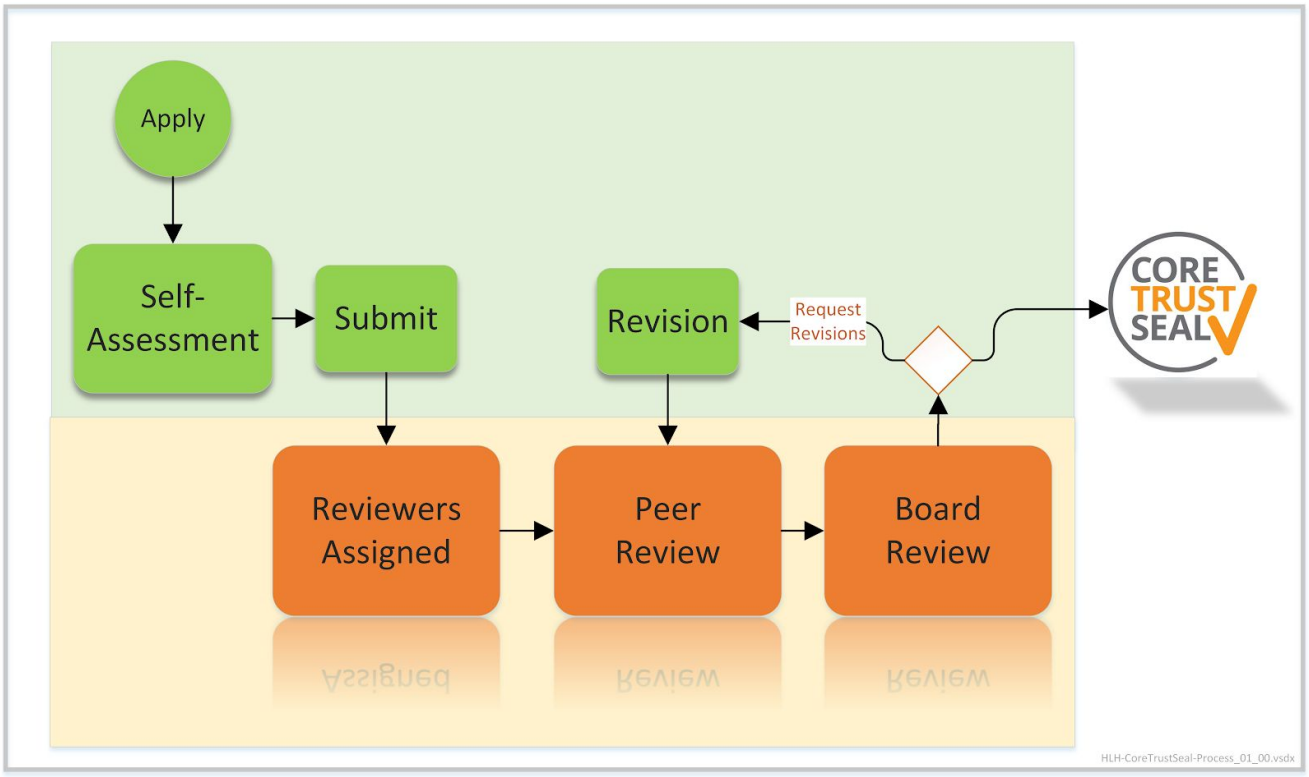


Diagram 8: ***CoreTrustSeal Process***

The diagram above presents the application in green (applicant) and the review process in orange (CoreTrustSeal Board). The self-assessment process supports defining a final assessment method which will result in agreed outcomes, including the defined 'status' of a repository, i.e. as a CoreTrustSeal

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

Trustworthy digital repository. The diagram below presents the addition of FAIRsFAIR repository support into the CoreTrustSeal process.

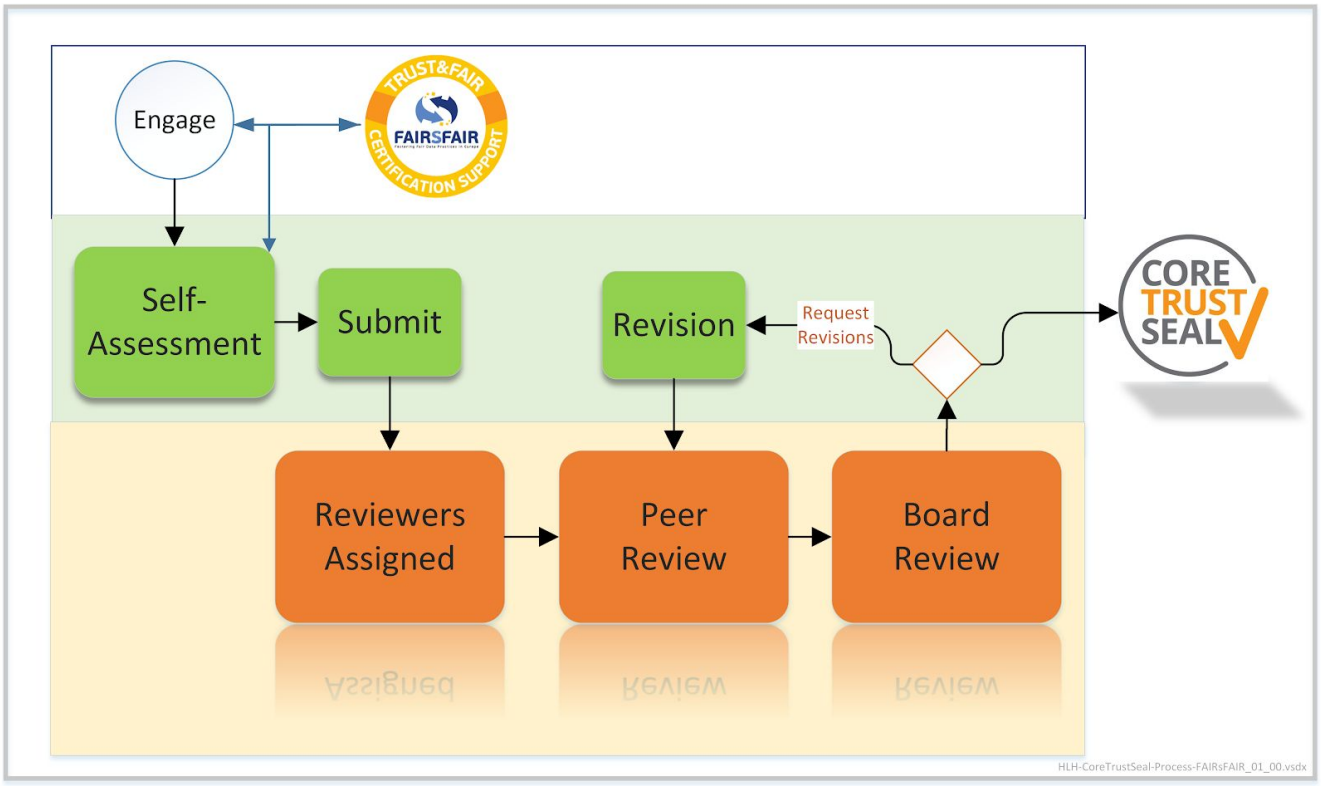


Diagram 9: ***FAIRsFAIR Project Repository Support***

This work takes place in parallel to efforts to test and 'badge' individual digital objects as 'FAIR'.

Repository support in FAIRsFAIR will enable applications for CoreTrustSeal which integrate evidence for FAIR enabling, but during this work there is no 'pass/fail' outcome within FAIRsFAIR or the formal process of FAIR enabled certification through CoreTrustSeal. Recommendations for integration are being shared and discussed with the CoreTrustSeal Board. The Board has provided a statement of support for this work (Appendix 3).

In designing evaluations and outcomes, we must also consider how to avoid unfairly penalising objects or repositories, especially in the design and testing phase of FAIR assessments, e.g. the protection of sensitive data should not lead to a lower score.

5.3.1.1. Certification and Badging

Beyond the design and implementation of indicators and tests for the FAIR principles, we will consider how best to recognise successful outcomes through formal certification and badging of FAIR entities. Certification and badging options have dependencies on the final structure of the approach and the different 'certification' actors that will be involved.

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

5.3.1.2. Change, Periodicity & Validity Terms

CoreTrustSeal repository certification lasts for three years. However, digital objects might change at any time. The period and terms under which a FAIR evaluation remains valid are important design considerations.

6. Open Issues for Integration

Our work to date has raised a number of issues, a selection of these are briefly outlined below. The issues will be considered in a future deliverable and further iterations of the CoreTrustSeal+FAIR approach. We would welcome feedback and input on each of these areas.

Iteration through Support and Wider Engagement

We are evaluating and testing a range of support approaches, including those used within the CESSDA Trust Support programme[25] also referenced by the SSHOC Project. Support and other engagement will help to define a flexible iteration schedule of design, implementation and evaluation throughout the project. Towards the end of the project, clear recommendations for the maintenance phase will be proposed.

Boundaries and Scope

Insourcing, outsourcing and complex partnerships can make repository boundaries hard to define. Complex, heterogeneous data collections can make it hard to define FAIRness at the repository level. The ability to clearly define the entity (object or organisation) under review is critical to any assessment, evaluation and certification process.

Registries

Registries will be a critical part of any future FAIR ecosystem. In addition to repository and object registries, FAIR principles and emerging indicators imply the need for other types. For example, do we need a clear registry of 'approved' PID systems, or disciplinary-specific data standard registries to help us evaluate 'rich' metadata?

Best, Minimal and Ideal Practices

The existence of standards like CoreTrustSeal, OAIS, ISO16363, ISO27001 and others does not mean there is always a community consensus on minimal levels of service quality and necessary supporting evidence. The CoreTrustSeal is the only current effort generating a publicly available body of work which could be used to support discussion on the often-used phrase 'best practices'. For formal assessment of object or repository characteristics, it is necessary to move from general assumptions of what 'best practice' means to SMART (specific, measurable, achievable, realistic, time-bound) objectives. We might also usefully differentiate between 'minimal practice' and 'ideal practice'. Some levels of practice might be

[25] https://doi.org/10.5281/zenodo.3621378 Overview of Support Approaches

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

defined purely from a 'technical perspective', e.g. a minimal number of data copies, while others will be dependent on local context including the needs of the data users.

Designated Community & other Users

"Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time".

Definition from the OAIS reference model as used by the CoreTrustSeal glossary[26].

For any real-world evaluation of an object, a repository or another FAIR entity, there must be a mixture of agreed practices and clear responsiveness to the changing needs of users. Whether this is a formally defined designated community, a broader mission to the public or a commercially driven approach based on supply (depositor) and demand (user). Some aspects of the evaluation must be based on whom a repository (or object, or service, etc.) is intended to serve.

We will seek more precise approaches to defining designated communities and agreement on expectations of how a repository should interact and respond to their needs.

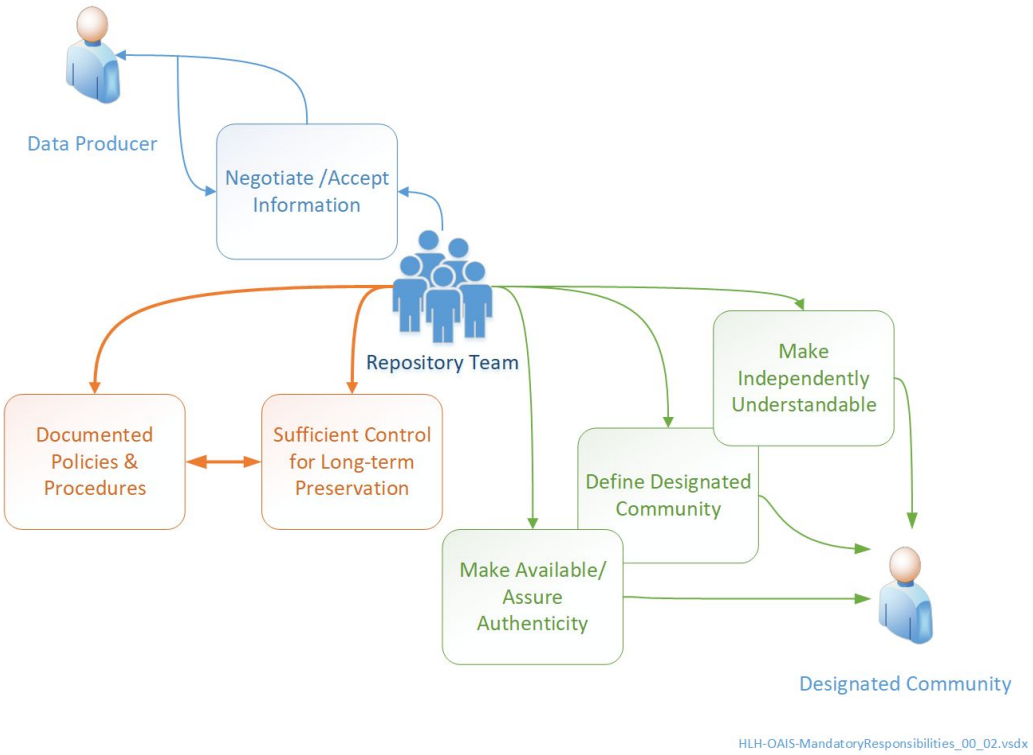


Diagram 10: ***OAIS Responsibilities Diagram***

[26] https://doi.org/10.5281/zenodo.3632563

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

The Full (FAIR) Data Lifecycle & Ecosystem

In line with the wider vision for FAIR, the FAIR-enabling repository work must integrate and align with a vision of the full FAIR data ecosystem and data lifecycle. This includes identifying how to align with work on research data management plans.

Non-(Meta) Data Artefacts as Evidence

Apart from a few cases where an entity (repository, object) is being directly inspected during a review there is always some dependency on evidence to support assessment. Evidence could range from mission statements, policies, procedures and workflows, to granular outcomes of fixity checks. This evidence is another type of 'digital object' generated as a result of running any infrastructure (people, processes, technology) which curates digital objects.

A key high-level indicator of maturity is the ability to design, implement, manage and change these evidence 'artefacts'. Without a business information management system, there will always be a risk of maintaining a consistent FAIRness level over time.

7. Conclusions and Next Steps

At this stage of the iterative process, we have a draft alignment between the FAIR Principles and the CoreTrustSeal Requirements. We have outlined the capability and maturity approach, which will be applied to the CoreTrustSeal+FAIR alignment. The responses from repositories to our high-level FAIR questions as they relate to repositories enabling FAIRness will help validate and improve the alignment. As we apply capability criteria to CoreTrustSeal+FAIR, we will address the calculation of overall repository maturity.

The overall goal is to develop a practical and sustainable approach for repositories to self-assess their current capability levels, identify target levels and define where they need to focus resources on improvement. Integration of these processes into operational practice will provide a common approach to assessing and evaluating a data repository's ability to enable FAIR data. The outcomes will be an overall improvement of repository practice and a pathway to certification.

A wide range of interactions and dependencies will influence this iterative work, including internal testing with supported repositories, external feedback and integration of ongoing developments. These include cooperation with the CoreTrustSeal Board and community. FAIRsFAIR supported repositories will be seeking to certify against the current version of the requirements, while the outcome of the project may recommend future directions for the structure, content and process of the CoreTrustSeal.

We are seeking comments, feedback and information about related efforts so that we can ensure cooperation, alignment and improvement of this crucial area of research data infrastructure.

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

8. Component Documents

8.1. Capability-Maturity Modeling and Landscape

This discussion paper provides an overview of the FAIRsFAIR[27] project approach to evaluating Capability Maturity Modelling for use alongside the alignment of the CoreTrustSeal Requirements[28] with the FAIR Data Principles[29].

https://doi.org/10.5281/zenodo.3862588

8.2. FAIR Principles: Baseline Comments

It is noticeable in various FAIR-related work that the same comments and questions related to the original Principles are repeatedly referenced. Rather than do the same thing for FAIRsFAIR WP4 we will retain the baseline issues and comments in this document and refer back to them periodically to see if they have been addressed either by our work or by others.

This text seeks to consider the issues around the FAIR Data Principles, particularly as they apply to the notion of a Trustworthy Digital Repository. Issues here must be answered (or at least acknowledged) for us to provide an aligned approach to FAIR-enabled Trustworthy Digital Repositories. We can progress without all of these questions being addressed, but clarifying them will ensure a better overall solution.

https://doi.org/10.5281/zenodo.3728131

8.3. FAIR Ecosystem Components: Vision

The primary focus of work package four in FAIRsFAIR is (trusted) repositories that enable the curation of (FAIR) objects. However, to be integrated into an operational European Open Science Cloud, a wider vision of FAIR ecosystem dependencies and interconnections is required. Data users and stewards of all kinds must be empowered to find, store and access data and metadata designed for interoperability and reuse. This draft presents a vision for the FAIR ecosystem components required to ensure FAIRness across the full data lifecycle.

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

8.4. CoreTrustSeal+FAIR Overview

This document represents the second alignment of CoreTrustSeal to FAIR requirements to inform repositories seeking to enable FAIR data. This version has been revised to include the latest version (v0.90) of the FAIR indicators developed by the RDA FAIR Data Maturity Working Group.

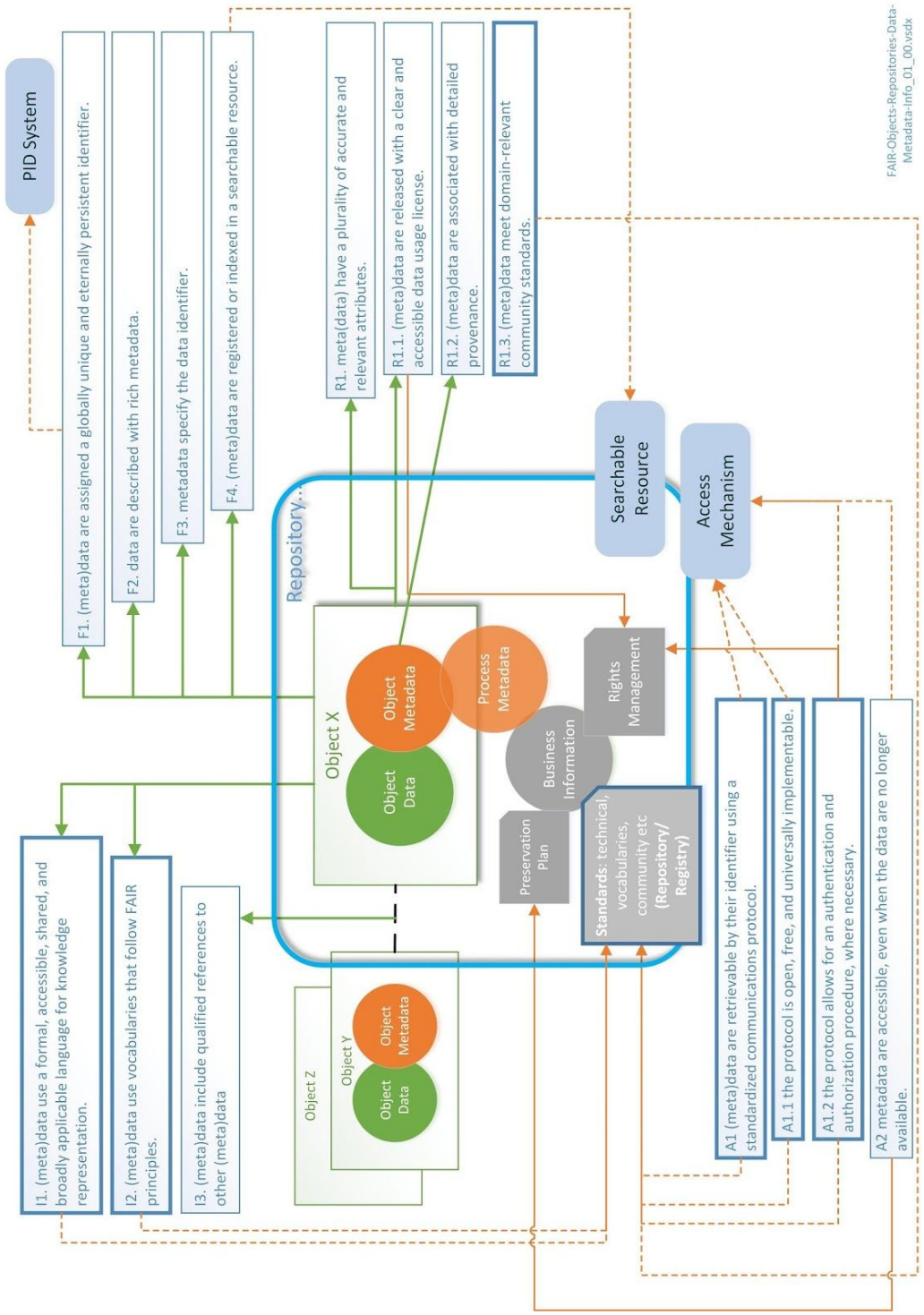
https://doi.org/10.5281/zenodo.3734896

[27] https://www.fairsfair.eu/
[28] https://www.coretrustseal.org/
[29] https://www.nature.com/articles/sdata201618

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

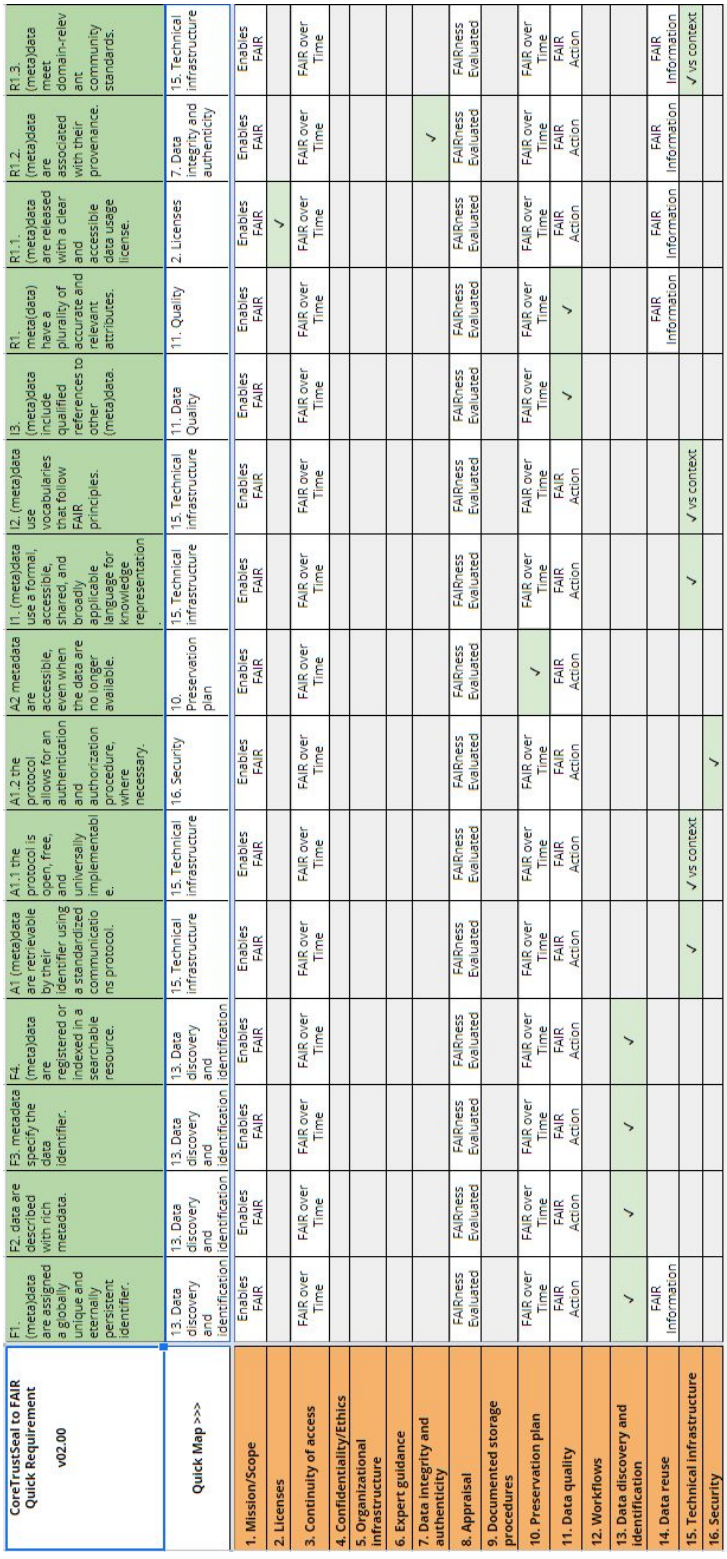
9. Appendix 1: FAIR Objects, Repositories, Dependencies

Objects and Repository characteristics are mapped to the FAIR Principles.



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

10. Appendix 2: CoreTrustSeal to FAIR: Quick Reference



CoreTrustSeal to FAIR Quick Requirement v02.00	F1. (meta)data are assigned a globally unique and eternally persistent identifier.	F2. data are described with rich metadata.	F3. metadata specify the data identifier.	F4. (meta)data are registered or indexed in a searchable resource.	A1 (meta)data are retrievable by their identifier using a standardized communicatio ns protocol.	A1.1 the protocol is open, free, and universally implementabl e.	A1.2 the protocol allows for an authentication and authorization procedure, where necessary.	A2. metadata are accessible, even when the data are no longer available.	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	I2. (meta)data use vocabularies that follow FAIR principles.	I3. (meta)data include qualified references to other (meta)data.	R1. meta(data) have a plurality of accurate and relevant attributes.	R1.1. (meta)data are released with a clear and accessible data usage license.	R1.2. (meta)data are associated with their provenance.	R1.3. (meta)data meet domain-relev ant community standards.
Quick Map >>>	13. Data discovery and identification	13. Data discovery and identification	13. Data discovery and identification	13. Data discovery and identification	15. Technical infrastructure	15. Technical infrastructure	16. Security	10. Preservation plan	15. Technical infrastructure	15. Technical infrastructure	11. Data Quality	11. Quality	2. Licenses	7. Data integrity and authenticity	15. Technical infrastructure
1. Mission/Scope	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR	Enables FAIR
2. Licenses													✓		
3. Continuity of access	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time
4. Confidentiality/Ethics															
5. Organizational infrastructure															
6. Expert guidance															
7. Data integrity and authenticity														✓	
8. Appraisal	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated	FAIRness Evaluated
9. Documented storage procedures															
10. Preservation plan	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	✓	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time	FAIR over Time
11. Data quality	FAIR Action	FAIR Action	FAIR Action	FAIR Action	FAIR Action	FAIR Action	FAIR Action	FAIR Action	FAIR Action	FAIR Action	✓	✓	FAIR Action	FAIR Action	FAIR Action
12. Workflows															
13. Data discovery and identification	✓	✓	✓	✓											
14. Data reuse	FAIR Information											FAIR Information	FAIR Information	FAIR Information	FAIR Information
15. Technical infrastructure					✓	✓ vs context			✓	✓ vs context					✓ vs context
16. Security							✓								

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

11. Appendix 3: CoreTrustSeal Board Statement

"FAIR data and other ongoing research data development have raised several key issues of relevance to CoreTrustSeal. We are actively engaging with FAIRsFAIR and a range of other FAIR-related projects and working groups. CoreTrustSeal-certified Trustworthy Data Repositories are vital components in enabling the realization of the Findable, Accessible, Interoperable, and Reusable (FAIR) Data Principles, both ensuring and enhancing the 'FAIRness' of data over the long term.

The mission for CoreTrustSeal endorsed by the Research Data Alliance and the wider community is to provide a single sustainable 'core' route for repository data service requirements and certification. The Board exists to manage and maintain that core route over time, and in response to community needs. As the FAIR Principles are clarified through indicators and evaluated through (ideally automated) tests against digital objects, CoreTrustSeal will continue to integrate 'core' best practices into the Requirements. We also recognise there may be more explicit FAIR requirements that may be elaborated around the foundation of the CoreTrustSeal. The CoreTrustSeal+FAIR work may be a case where we can integrate a FAIR-enabling assessment into the CoreTrustSeal process.

The CoreTrustSeal Board will continue to follow and engage in the work carried out by FAIRsFAIR and other FAIR-related initiatives around the world to ensure that CoreTrustSeal certification continues to address community needs for core-level certification."

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