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

CoreTrustSeal^{+}FAIR Overview

Work Package	WP4 FAIR-Certification
Lead Author (Org)	Hervé L'Hours (UKDA)
Contributing Author(s) (Org)	Hervé L'Hours (UKDA), Ilona von Stein, Frans Huigen, Mustapha Mokrane, Jerry de Vries, Linas Cepinskas (DANS), Anusuriya Devaraju, Robert Huber (UniHB), Joy Davidson, Patricia Herterich (DCC)
Due Date	31.08.2020
Date	31.08.2020
Version	3.0
DOI	https://doi.org/10.5281/zenodo.4003630

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)

Table of contents

CoreTrustSeal+FAIR 3

1. Introduction 4

2. CoreTrustSeal+FAIR: High-Level Discussion Points 7

Findable 8

Accessible. 8

Interoperable 8

Reusable 8

3. Additional Context: High Level Repository Capabilities **9**

3.1 Context: Stakeholder Ecosystem & Standards 9

3.2 Context. Business Information Management: Evidence 9

3.3 Context. Business Systems: Managing Change 10

3.4 Context. Data & Collection Context 10

4. CoreTrustSeal Requirements & FAIR (towards CoreTrustSeal+FAIR) **11**

Background Information 11

R0. Context. 11

Organisational Infrastructure 12

R1. Mission/Scope 12

R2. Licenses 12

R3. Continuity of access 13

R4. Confidentiality/Ethics 14

R5. Organizational infrastructure 14

R6. Expert guidance 14

Digital Object Management 15

R7. Data integrity and authenticity 15

R8. Appraisal 16

R9. Documented storage procedures 16

R10. Preservation plan 17

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

R11. Data quality 17

R12. Workflows 19

R13. Data discovery and identification 19

R14. Data reuse 21

Technology 22

R15. Technical infrastructure 22

R16. Security 25

Appendix 1: FAIR Principles to CoreTrustSeal Alignment 26

Appendix 2: CoreTrustSeal to FAIR Mapping: Quick Reference 27

CoreTrustSeal+FAIR

Defining an object as FAIR (Findable, Accessible, Interoperable, Reusable) is to a great degree dependent on its curation context. Maintaining FAIRness over the long term depends on the preservation of an object. The CoreTrustSeal+FAIR work examines the alignment between the FAIR Data Principles with the CoreTrustSeal Trustworthy Digital Repository Requirements and considers how increased tiers of FAIRness and Trust can be described through capability/maturity levels.

This approach will be tested within the FAIRsFAIR project. Recommendations for integration are being shared and discussed with the CoreTrustSeal Board. The Board has provided a statement of support for this work, but no direct alignment with the CoreTrustSeal or its processes is currently in place. In the project timeframe, there is no formal process of FAIR enabled certification through CoreTrustSeal. This would require adoption through the periodic CoreTrustSeal community review of requirements and processes.

We are elaborating more specific requirements around the ‘Core’ of the CoreTrustSeal. An alignment between object FAIRness and trustworthy repository standards provides benefits to funders, depositors, repository data services and their users either with, or without full formal 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

1. Introduction

This document represents the third alignment of CoreTrustSeal to FAIR requirements to inform repositories seeking to enable FAIR data.

"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 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."[1]

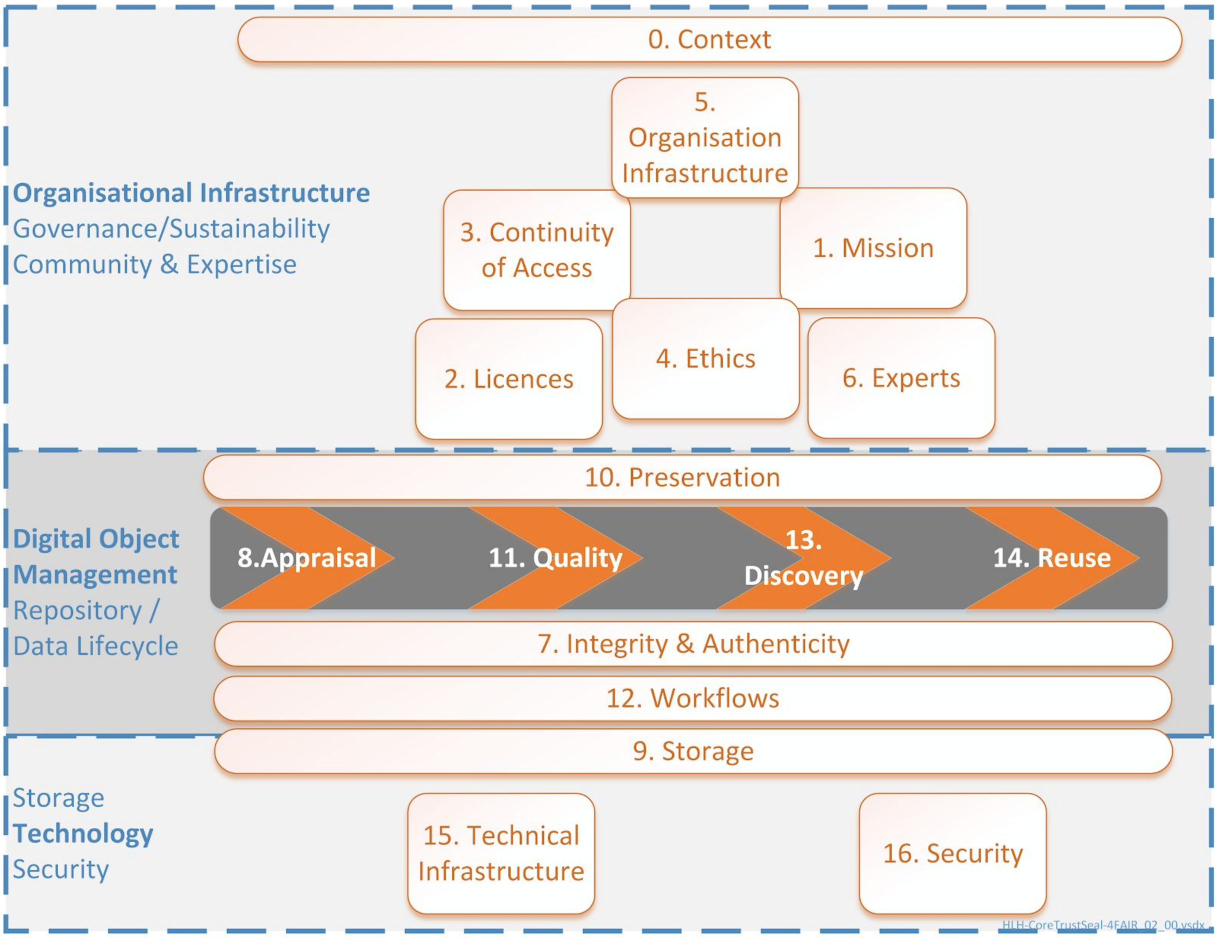

0. Context
5. Organisation Infrastructure
Organisational Infrastructure
Governance/Sustainability
Community & Expertise
3. Continuity of Access
1. Mission
2. Licences
4. Ethics
6. Experts
10. Preservation
Digital Object Management
Repository / Data Lifecycle
8.Appraisal
11. Quality
13. Discovery
14. Reuse
7. Integrity & Authenticity
12. Workflows
9. Storage
Storage
Technology
Security
15. Technical Infrastructure
16. Security
HLH-CoreTrustSeal-4FAIR_02_00.vsdx

Diagram 1: ***CoreTrustSeal Requirements in Brief***

The ten repositories receiving support to achieve CoreTrustSeal through the FAIRsFAIR project responded to the high level questions and discussion points raised. Those responses will be used to guide FAIR-related aspects of the support process.

[1] D4.2 Repository Certification Mechanism

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 supported repositories have now completed their first self-assessments against the CoreTrustSeal Requirements (without FAIR elements). The support process will provide guidance on how best to improve self-assessment statements and supporting evidence. The next iteration of self-assessment will be used to re-integrate FAIR into the requirements evaluation and consider the relationship between capability, evidence and ultimately overall organisational (repository) maturity.”[2]

The FAIRsFAIR Data Object Assessment Metrics[3] are fifteen minimum viable metrics proposed by FAIRsFAIR for the systematic assessment of FAIR data objects. They were based on indicators proposed by the RDA FAIR Data Maturity Model Working Group[4], on the WDS/RDA Assessment of Data Fitness for Use checklist[5], and on prior work conducted by project partners such as FAIRdat[6] and FAIREnough[7].

The CoreTrustSeal Requirements have been mapped to the FAIR Principles, associated Research Data Alliance Working Group indicators and the subset of FAIRsFAIR Data Object Assessment Metrics designed to test against the indicators. In this iteration the Principle to Requirements crosswalk remains stable (see Appendix 1). The table in Appendix 2 has been revised to include ‘B’ level indicators where feedback indicates that the Principle, Indicator or Metric has ‘some’ association with a CoreTrustSeal Requirement. Our primary focus remains on ‘A’ level indicators as self-assessments against CoreTrustSeal ‘plus’ FAIR need a single, clear location to provide evidence. Appendix 2 mappings which include “vs Context'' indicates that evidence could be provided as part of the Context information (R.0) in a future version of the CoreTrustSeal Requirements.

Earlier consultation proposed an approach that used the CMMI 2.0[8] capability maturity tiers.

CMMI Levels.

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

For this iteration of the CoreTrustSeal+FAIR overview we propose a standard target capability ‘level of 3: Defined’ for each Requirement. This (initially self-assessed) capability level exists alongside the relevant CoreTrustSeal Compliance level.

[2] M4.2 Draft Maturity Model CoreTrustSeal Requirements
[3] FAIRsFAIR Data Object Assessment Metrics
[4] FAIR Data Maturity Model
[5] WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations
[6] https://docs.google.com/forms/d/e/1FAIpQLSd8_pd2r2SnjCVfCC3CHhEUHZzv2MTRC3RTh0S2YTvbVJj87Q/viewform
[7] https://docs.google.com/forms/d/e/1FAIpQLSf7t1Z9IOBoj5GgWqik8KnhtH3B819Ch6lD5KuAz7yn0I0Opw/viewform
[8] https://cmmiinstitute.com/cmmi-v2-0-model-at-a-glance

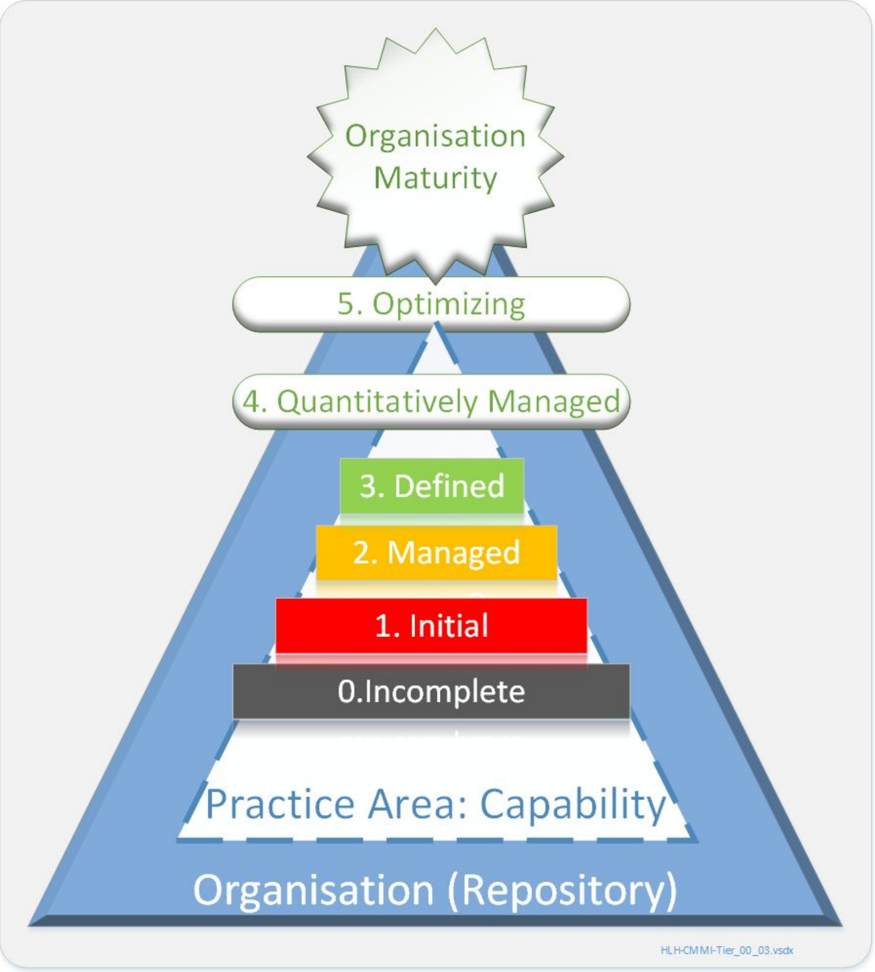


Diagram 2: ***Tiered Capability/Maturity***[9]

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

The CoreTrustSeal Guidance[10] notes that Compliance Levels of 1 or 2 are not sufficient for a successful application. Certification may be granted if some Requirements are in the implementation phase (3).

The next step will be to examine the outcomes of automated or manual tests against FAIR indicators alongside the evidence sought for each CoreTrustSeal Requirement and to consider how

[9] D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures

[10] CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022

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

they indicate levels of capability and maturity. Further details of capability/maturity assessments and their applications are presented in the FAIRsFAIR discussion paper Landscape of Capability Maturity Modelling.[11]

The scope of the CoreTrustSeal+FAIR mapping may be influenced by a number of ongoing activities both within and beyond the FAIRsFAIR project work. These include FAIRsFAIR *D3.3 Policy Enhancement Recommendations*[12] and *D3.4 Recommendations on practice to support FAIR data principles*[13]. We will consider the degree to which policy and practice recommendations can or should be integrated into the CoreTrustSeal+FAIR and whether they can inform more detailed statements on what we expect at each capability level.

Along with direct feedback from ten supported repositories FAIRsFAIR is developing a wider range of repository support approaches. The text of, and emerging feedback to *M3.5 Description of Transition Support Programme for Repositories*[14] raises a number of issues relevant to CoreTrustSeal+FAIR including clarification of standards in place at a repository, long term preservation definitions, repository types and data types. These also reflect some of the issues implied by the recent CoreTrustSeal consultation[15] on specialist (e.g. domain/disciplinary) repositories vs more generalist repositories and the outsourcing of repository functions to technical repository service providers (TRSP)

Additional information about the design principles and context around these mappings are presented in a separate FAIRsFAIR Milestone document[16].

2. CoreTrustSeal+FAIR: High-Level Discussion Points

The clarification of various FAIR principles and the best approach to FAIR indicators and tests are still in progress. The FAIRsFAIR approach will evolve along with those clarifications. Questions below are a high-level starting point for considering repository/FAIR context. These will evolve and be further integrated into CoreTrustSeal+FAIR as they are clarified, contextualised and as indicators are defined and test processes agreed. This section provides some CoreTrustSeal+FAIR 'conversation starters' for those repositories beginning the CoreTrustSeal+FAIR journey.

This section references each FAIR principle, but does not address the indicators[17]. Indicators are mapped in section 4 at the Requirements level.

[11] CoreTrustSeal+FAIR Landscape of Capability Maturity Modeling - A FAIRsFAIR Discussion Paper

[12] D3.3 Policy Enhancement Recommendations

[13] D3.4 Recommendations on practice to support FAIR data principles

[14] M3.5 Description of Transition Support Programme for Repositories

[15] https://www.coretrustseal.org/why-certification/specialists-generalists-technical-repository-service-providers/

[16] M4.2 Draft Maturity Model CoreTrustSeal Requirements

[17] 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

Findable

*"**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.*

Question: What persistent identifier system do you use? Are any of your objects not persistently identified? Which search interfaces provide access to your objects? Which types of users (human or machine) are you targeting by using those interfaces? What metadata standards are used to support resource discovery? Are any of your objects not available in a resource discovery system?

Response:

Accessible.

*"**A1.** (meta)data are retrievable by their identifier using a standardized communications protocol. **A1.1.** the protocol is open, free, and universally implementable. **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."*

Question: What different levels of data access do you offer for your objects? By which methods and technologies do your users' retrieve objects? When objects are removed from your collections do their metadata remain available?

Response:

Interoperable

*"**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".*

Question: How do you understand the term interoperability in the context of your (meta) data and your users? What formats and schemas do you use for your (meta)data. Which vocabularies do you use and how are they managed? How do you build links within your (meta)data collections and out to (meta)data in other collections?

Response:

Reusable

*"**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-relevant community standards."*

Question: How do you understand the term reusable in the context of your (meta) data and your users? What licenses do you apply and communicate to users? How do you document changes to

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 (meta)data? What is your version model? What meta(data) standards do you use? How are these standards defined and managed?

Response:

3. Additional Context: High Level Repository Capabilities

The additional context items below may be integrated into FAIRsFAIR self-assessments for CoreTrustSeal 'R0: Context' once refined and agreed upon.

3.1 Context: Stakeholder Ecosystem & Standards

Technical standards fit into R15 Technical Infrastructure. But whether a standard is 'technical' might be open to debate.

Question: What legal, ethical or other 'non-technical' standards apply to your repository services and to making data FAIR?

Response:

3.2 Context. Business Information Management: Evidence

A key dependency for any self-assessment and any programme of operational change is a collection of business information and supporting processes that provide evidence for repository practice.

Question: Briefly describe the process for developing, implementing, reviewing and applying policies and procedures in your organisation. Do you integrate data management plans (DMP) into your processes, if so how?

Response:

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

3.3 Context. Business Systems: Managing Change

The people, processes and technologies which make up data infrastructure must be capable of managing change over time in response to new circumstances and to improve services.

Question: Briefly describe any change management processes and procedures in place in your organisation.

Response:

3.4 Context. Data & Collection Context

A broad understanding of the objects being curated and how they are grouped into collections is important context for CoreTrustSeal+FAIR.

Question: Briefly describe the characteristics of the data you curate that you consider important. Describe how the data is divided up into different collections and why (different data types, different audiences etc). What is your approach to describing the digital objects you hold in terms of their data, metadata and documentation content? How are these digital objects supported by other metadata such as technical, administrative or preservation metadata?

Response:

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

4. CoreTrustSeal Requirements & FAIR (towards CoreTrustSeal+FAIR)

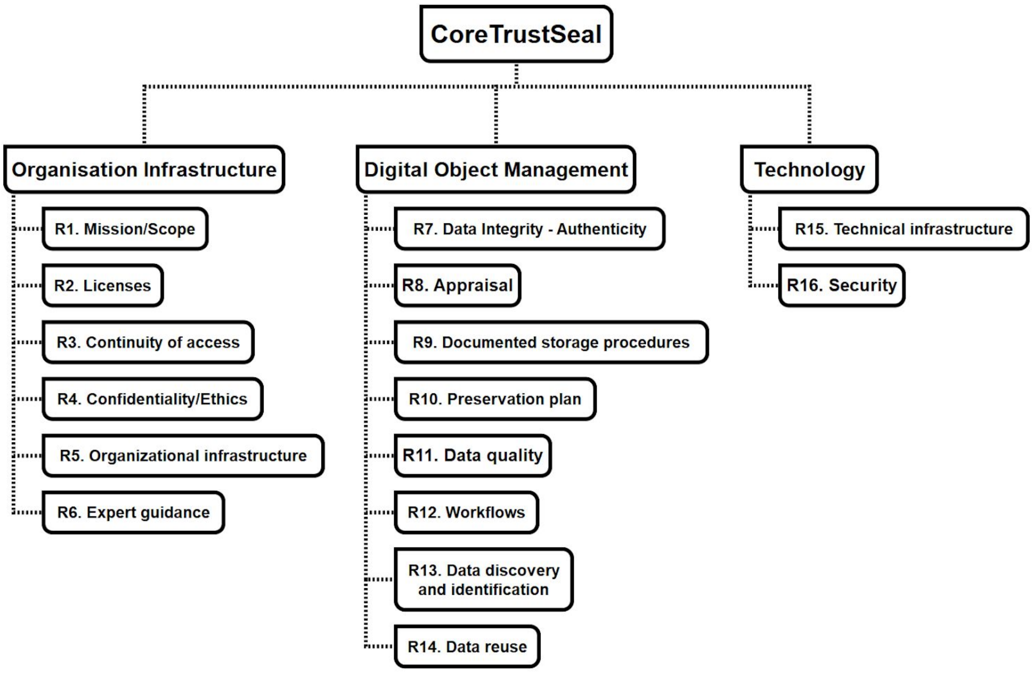

CoreTrustSeal
Organisation Infrastructure
R1. Mission/Scope
R2. Licenses
R3. Continuity of access
R4. Confidentiality/Ethics
R5. Organizational infrastructure
R6. Expert guidance
Digital Object Management
R7. Data Integrity - Authenticity
R8. Appraisal
R9. Documented storage procedures
R10. Preservation plan
R11. Data quality
R12. Workflows
R13. Data discovery and identification
R14. Data reuse
Technology
R15. Technical infrastructure
R16. Security

Diagram 3: ***CoreTrustSeal Requirements***

See appendices 1 and 2 for a diagram and an alignment table mapping FAIR to CoreTrustSeal.

Background Information

R0. Context.

R0. Context. Repository Type

Questions: Do these repository types apply to your repository? Is there anything about your repository type which influences how you enable FAIR data?

Response:

The following items from CoreTrustSeal have no specific mappings to FAIR principles at this stage:

R0. Context. Brief Description of Repository
R0. Context. Brief Description of the Designated Community
R0. Context. Level of Curation Performed
R0. Context. Insource/Outsource Partners
R0. Context. Summary of Significant Changes

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

R0. Context. Other Relevant Information

NB: Once agreed the additional items above under "Additional Context" may be integrated here.

Organisational Infrastructure

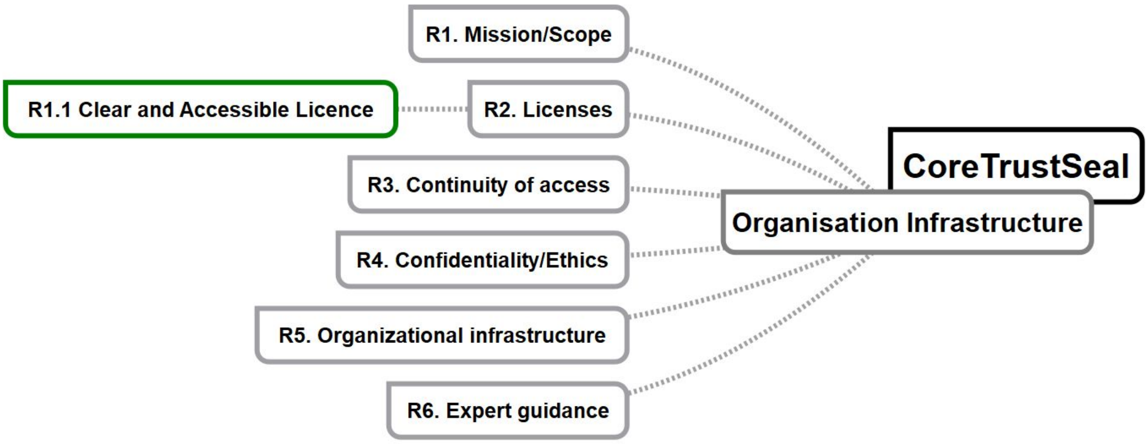

R1. Mission/Scope
R1.1 Clear and Accessible Licence
R2. Licenses
R3. Continuity of access
R4. Confidentiality/Ethics
R5. Organizational infrastructure
R6. Expert guidance
CoreTrustSeal
Organisation Infrastructure

*Diagram 4: **CoreTrustSeal Requirements: Organisational Infrastructure***

R1. Mission/Scope

No 'A' level mapping to FAIR.

Question: Does your evidence related to 'mission' specifically reference findability, accessibility, interoperability or re-usability? Should it?

Response:

Discussion: Access is assumed to be a primary mission of a TDR, but it does not have its own Requirement in CoreTrustSeal. Do we need to add anything specific to ensure +FAIR?

Comments:

R2. Licenses

Note: Principle '*A1.2 the protocol allows for an authentication and authorization procedure, where necessary*' depends on criteria set by licences, but A1.2 is mapped to R16 Security. Licences may depend on confidentiality and ethical issues are addressed under R4.

***Principle**: R1.1. (meta)data are released with a clear and accessible data usage license.*

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

FAIRsFAIR Metric:

FsF-R1.1-01M Metadata includes license information under which data can be reused.

RDA Indicators:

- RDA-R1.1-01M Metadata includes information about the licence under which the data can be reused (Essential)
- RDA-R1.1-02M Metadata refers to a standard reuse licence (Important)
- RDA-R1.1-03M Metadata refers to a machine-understandable reuse licence (Important)

Question: How do you approach rights management including deposit and access licence management and intellectual property rights? What levels of access conditions are applied to your objects? What metadata standard is used for rights information?

Response:

Discussion: Do all of the indicators defined need to be met to ensure +FAIR?

Comments:

Note: 'Principle: A1 (meta)data are retrievable by their identifier using a standardized communications protocol' has an indicator: "RDA-A1-01M "Metadata contains information to enable the user to get access to the data (Important) ". We have mapped Principle A1 to R15 Technical Infrastructure even though the indicator also aligns with Licences as regards access conditions. This is an example where we need to make practical choices about mapping CoreTrustSeal to FAIR and R2 is a 'B' level alignment.

Comments:

R3. Continuity of access

No 'A' level mapping to FAIR

Continuity of access reduces the risks to FAIR Data by ensuring it is cared for in a repository that addresses business continuity, disaster recovery and succession planning.

This requirement is primarily repository rather than object focussed. "*A2 metadata are accessible, even when the data are no longer available*" has some relevance here but is addressed under R10. Preservation. "*R1.1. (meta)data are released with a clear and accessible data usage license*" (covered under R2 Licences) is a dependency for succession planning in R3.

Discussion: Does the assessment need to differentiate between different aspects of Continuity of Access? E.g. Disaster Recovery, Business Continuity & Succession Planning? Do we need to add anything else here to ensure +FAIR?

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

R4. Confidentiality/Ethics

No ‘A’ level mapping to FAIR

This requirement includes a focus on practices which manage sensitive data, personal data and disclosure risk. “*A1.2 the protocol allows for an authentication and authorization procedure, where necessary*” is relevant here, but this is addressed under R16. Security. An understanding of Confidentiality/Ethics around an object is a dependency for ‘‘*R1.1. (meta)data are released with a clear and accessible data usage license*” but this is addressed under R2 Licences.

Discussion: Do we need to add anything else here to ensure +FAIR?

Comments:

R5. Organizational infrastructure

No ‘A’ level mapping to FAIR

This requirement is less focussed on objects and primarily focussed on repository characteristics.

Discussion: It is not immediately clear how such a potentially complex area as organisational infrastructure capability, including resources, governance and skills, can be quickly assessed at a ‘core’ level without being made more granular . There is an open question as to whether some of the additional context questions for FAIR enabling might be placed under R5: Organisational Infrastructure. Do we need to add anything else here to ensure +FAIR?

Comments:

R6. Expert guidance

No ‘A’ level mapping to FAIR

External expert guidance may be a dependency for defining relevant context for the data collection, e.g. “*R1.3. (meta)data meet domain-relevant community standards*” but this would be assured as part of Data Quality (R11) and the standards listed under R15 Technical Infrastructure.

Discussion: This is a very broad area to apply a capability maturity level. Expert Guidance might support, for example, the selection of an appropriate domain standard. Would a format registry be considered ‘expert guidance’? Do we need to add anything else here to ensure +FAIR?

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

Digital Object Management

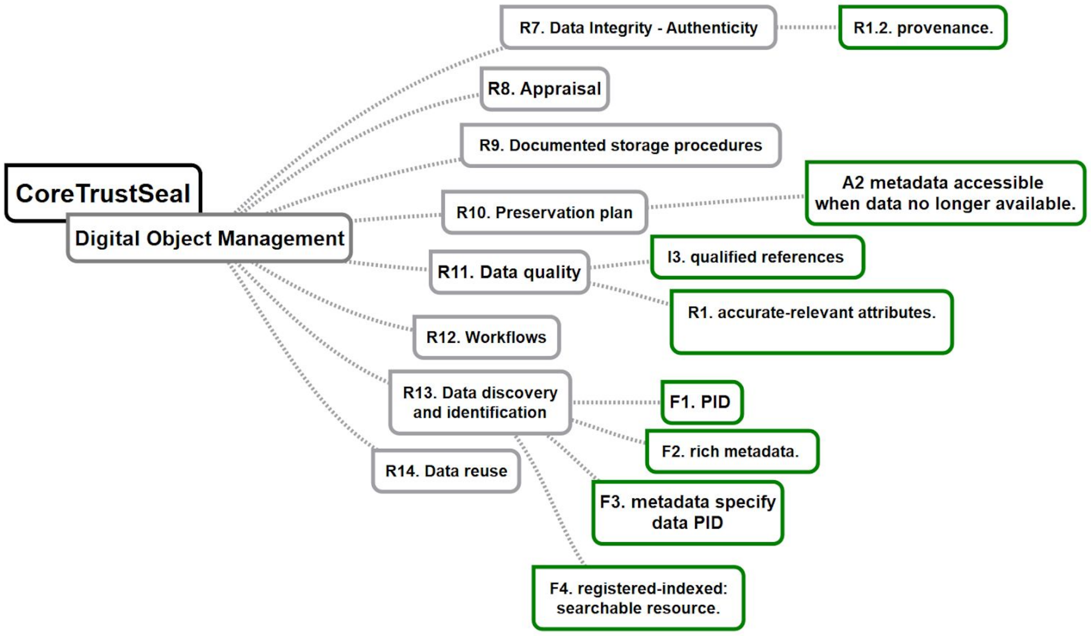

CoreTrustSeal
Digital Object Management
R7. Data Integrity - Authenticity
R1.2. provenance.
R8. Appraisal
R9. Documented storage procedures
R10. Preservation plan
A2 metadata accessible when data no longer available.
R11. Data quality
I3. qualified references
R1. accurate-relevant attributes.
R12. Workflows
R13. Data discovery and identification
F1. PID
F2. rich metadata.
F3. metadata specify data PID
F4. registered-indexed: searchable resource.
R14. Data reuse

Diagram 5: ***CoreTrustSeal Requirements: Digital Object Management***

R7. Data integrity and authenticity

Integrity

Data integrity is not directly addressed by the FAIR Principles.

Authenticity

Principle: *R1.2. (meta)data are associated with their provenance.*

FAIRsFAIR Metric

FsF-R1.2-01M Metadata includes provenance information about data creation or generation.

RDA Indicators:

- RDA-R1.2-01M Metadata includes provenance information according to community-specific standards (Important)
- RDA-R1.2-02M Metadata includes provenance information according to a cross-community language (Useful)

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 FAIR Principles reference provenance as part of ‘Reuse’ but this principle maps to R7 (Authenticity) in CoreTrustSeal. The focus of the indicators here is on ‘standards’ which should either form part of the technical standards under R15 or as part of higher level context (see *Context: Stakeholder Ecosystem & Standards above*).

Question: What information about integrity measures do you take at the point of deposit, during curation and for data at the point of access? What provenance standards in which ‘cross community languages’ do you have in place? How are these applied and communicated to users?

Response:

Discussion: In this case it seems unlikely that both integrity and provenance would always be assigned the same capability level. Do we need to add anything else here to ensure +FAIR?

Comments:

R8. Appraisal

No ‘A’ level mapping to FAIR.

The level of FAIRness of an object and the level to which the FAIR Principles can be applied to a digital object should be evaluated during the Appraisal process. Any FAIR Principles that an object does not comply with at the point of deposit should be addressed during curation (R11. Data Quality). Any FAIR Principles which cannot be met should be communicated and explained to users at the point of ReUse (R14).

Question: What information related to the FAIRness of objects do you collect at the point of deposit. Could a lack of FAIRness be a reason for refusing a deposit? Do you have data which cannot be made FAIR? Why?

Response:

Discussion: How can FAIRness be best evaluated at the point of appraisal/deposit?

Comments:

R9. Documented storage procedures

No ‘A’ level mapping to FAIR.

The FAIR Principles do not directly address data storage.

Discussion: This seems like a case where community-agreed levels of capability could be defined. Do we need to add anything else here to ensure +FAIR?

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

R10. Preservation plan

All of the FAIR data principles reflect common repository practices to support long term preservation access. The addition of trustworthy digital repository practices to the FAIR principles ensures that the FAIR status of an object is more than a 'snapshot' in time. CoreTrustSeal^{+}FAIR helps ensure FAIRness over time.

***Principle**: A2 metadata are accessible, even when the data are no longer available.*

FAIRsFAIR Metric

FsF-A2-01M	Metadata remains available, even if the data is no longer available.

RDA Indicator

- RDA-A2-01M Metadata is guaranteed to remain available after data is no longer available (Essential)

Principle A2 and its indicator are an explicit requirement that metadata is preserved. But this is also associated with standard practice for persistent identifier management (R13 Data Discovery and Identification).

Question: Does your preservation plan make it explicit that metadata must remain available even when an object is removed from your repository?

Response:

Discussion: What level of additional detail on FAIRness is required to demonstrate preservation of FAIR data characteristics?

Comments:

R11. Data quality

'*R1.3. (meta)data meet domain-relevant community standards*' depends on data quality steps to ensure standards compliance. But standards are addressed under R15 Technical Infrastructure in CoreTrustSeal, or this could form part of *Stakeholder Ecosystem and Standards* under *Additional Context* (above).

Any lack of FAIRness identified during Appraisal (R8) should be addressed as part of curation to ensure quality. Quality standards, including FAIRness (or lack of FAIRness) should be communicated to users at the point of Re-use (R14).

Question: What steps does your repository take to ensure FAIRness during curation for quality?

Response:

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

***Principle**: I3. (meta)data include qualified references to other (meta)data.*

FAIRsFAIR Metric

FsF-I3-01M	Metadata includes links between the data and its related entities.

RDA Indicators

- RDA-I3-01M Metadata includes references to other metadata (Important)
- RDA-I3-01D Data includes references to other data (useful)
- RDA-I3-02M Metadata includes references to other data (Useful)
- RDA-I3-02D Data includes qualified references to other data (Useful)
- RDA-I3-03M Metadata includes qualified references to other metadata (Important)
- RDA-I3-04M Metadata include qualified references to other data (Useful)

Question: How does your (meta) data provide links to other (meta) data and why?

Response:

Discussion: I3 presents a challenge for mapping to CoreTrustSeal. Though a rich network of linked data and metadata objects can be very valuable the principle and indicators are very broad. We have mapped to R11. Data quality as this principle suggests curation to comply with a clearly articulated object model. There is some question over whether this would depend on R15 Technical Infrastructure standards or if this is a less technical type of standardisation.

Comments:

***Principle**: R1. meta(data) have a plurality of accurate and relevant attributes.*

FAIRsFAIR Metric

FsF-R1-01MD	Metadata specifies the content of the data.

RDA Indicator:

- RDA-R1-01M Plurality of accurate and relevant attributes are provided to allow reuse (Essential)

There is some overlap between the ‘Findability’ focussed “*F2. data are described with rich metadata*” and “*R1. meta(data) have a plurality of accurate and relevant attributes*.”

Question: How do you identify whether metadata is sufficient for reuse by your users?

Response:

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

Discussion: How can FAIRness be best evaluated during data quality assurance/curation steps?

Comments:

R12. Workflows

No ‘A’ level mapping to FAIR

Workflows are not directly addressed by the FAIR Principles. Though defined, managed and recorded workflows within the repository are dependencies for provenance related to the repository portion of the data lifecycle (*R1.2. (meta)data are associated with their provenance*).

Question: How do you develop, implement and manage change to repository workflows?

Response:

Discussion: Workflows to manage evidence artefacts (mission statements, licences, business continuity plans, legal ethical compliance, storage procedures, governance information, preservation plans, technical infrastructure and security) and activities (appraisal, quality assurance, re-use etc) may be evaluated at different capability levels. Workflows are also a dependency for overall organisational maturity. Are there elements of FAIRness that should be explicitly addressed in workflows?

Comments:

R13. Data discovery and identification

CoreTrustSeal R13 maps closely to the Findable Principles.

There is also an association between discovery, access and reuse. The provision of ‘Access’ is assumed to be part of the trustworthy digital repository mission (R1) so it is implied throughout CoreTrustSeal rather than addressed separately. But “*A1 (meta)data are retrievable by their identifier using a standardized communications protocol*” is mapped to R15 Technical Infrastructure.

***Principle**: F1. (meta)data are assigned a globally unique and eternally persistent identifier.*

FAIRsFAIR Metric

FsF-F1-01D	Data is assigned a globally unique identifier.
FsF-F1-02D	Data is assigned a 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

RDA Indicators:

F1	RDA-F1-01M	Metadata is identified by a persistent identifier (Essential)
F1	RDA-F1-01D	Data is identified by a persistent identifier (Essential)
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier (Essential)
F1	RDA-F1-02D	Data is identified by a globally unique identifier (Essential)

Question: Are all of the data in your collection assigned a PID? If not, why not?

Response:

Principle*: F2. data are described with rich metadata.*

FAIRsFAIR Metric

FsF-F2-01M Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.

RDA Indicator:

F2	RDA-F2-01M	Rich metadata is provided to allow discovery	(Essential)

Question: What metadata do your users need to support resource discovery? Does this metadata follow domain/discipline-specific standards? Which ones?

Response:

Principle*: F3. metadata specify the data identifier.*

FAIRsFAIR Metric

FsF-F3-01M Metadata includes the identifier of the data it describes.

Indicator:

F3	RDA-F3-01M	Metadata includes the identifier for the data	(Essential)

Question: Does the metadata for all of your objects include the identifier for the data it describes? If not, why not?

Response:

Principle*: F4. (meta)data are registered or indexed in a searchable resource.*

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

FAIRsFAIR Metric

FsF-F4-01M Metadata is offered in such a way that it can be retrieved by machines.

RDA Indicator:

- F4 RDA-F4-01M Metadata is offered in such a way that it can be harvested and indexed (Essential)

Question: Through which systems can your users discover your resources? Do these systems follow domain/disciplinary standards? If so, which? If not, why not?

Response:

Discussion: it is notable that Principles F1 to F3 relate to the characteristics of a (meta) data object. As collections may be heterogeneous there is an argument for asking additional questions as part of '*Context. Data & Collection Context' above*. The overall 'profile' of the repository collection will impact all of the CoreTrustSeal and FAIR assessment items.

Comments:

R14. Data reuse

R14. Data Reuse is the intuitive mapping for the R in FAIR. But the Principles themselves are more granular, as are the potential metrics and tests. Aspects of FAIR Re-use are addressed more broadly elsewhere under CoreTrustSeal.

No 'A' level mapping to FAIR

FAIRness is assured through curation actions associated with R11. Data Quality. The FAIRness (or otherwise) of the (meta)data should be communicated to users at the point of reuse.

> '*Principle: R1. meta(data) have a plurality of accurate and relevant attributes*' is critical for re-use but has been mapped to R11. Data Quality as that is where the curation processes to ensure these characteristics take place.
>
> '*Principle: R1.1. (meta)data are released with a clear and accessible data usage license*'. Under FAIR this is part of Reuse, but within CoreTrustSeal it must form part of the overall rights management (R2 Licences) above.
>
> '*Principle: R1.2. (meta)data are associated with their provenance*'. Provenance is vital for re-use, but within CoreTrustSeal it falls under overall data integrity and authenticity (R7) above.
>
> '*Principle: R1.3. (meta)data meet domain-relevant community standards*'. The focus here is on ' standards' which should either form part of the technical standards under R15 (added below) or as part of higher level context (see Context: Stakeholder Ecosystem & Standards above)

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

Technology

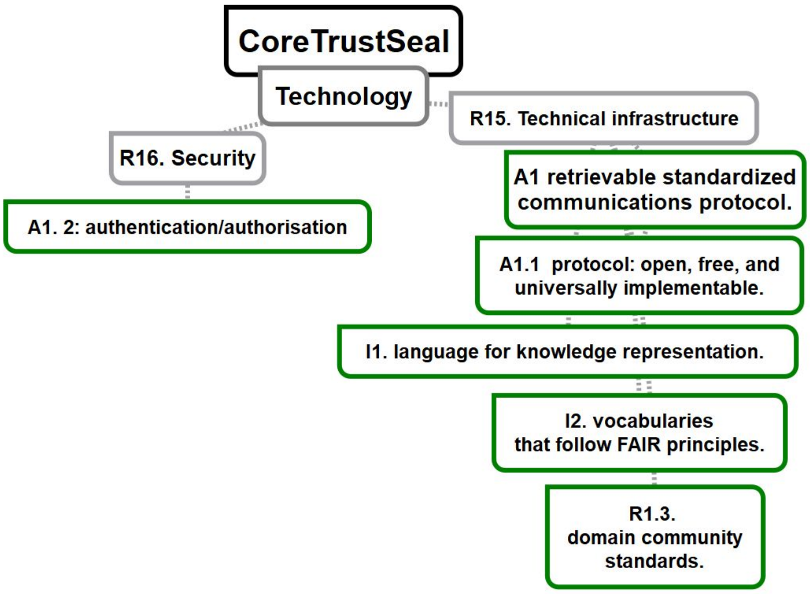

CoreTrustSeal
Technology
R15. Technical infrastructure
R16. Security
A1. 2: authentication/authorisation
A1 retrievable standardized communications protocol.
A1.1 protocol: open, free, and universally implementable.
I1. language for knowledge representation.
I2. vocabularies that follow FAIR principles.
R1.3. domain community standards.

*Diagram 6: **CoreTrustSeal Requirements: Technology***

R15. Technical infrastructure

***Principle**: "A1 (meta)data are retrievable by their identifier using a standardized communications protocol."*

FAIRsFAIR Metric

FsF-A1-01M Metadata contains access level and access conditions of the data.

This FAIRsFAIR metric also has a 'B' Level association with R2. Licences

RDA Indicator:

- RDA-A1-01M Metadata contains information to enable the user to get access to the data (Important)
- RDA-A1-02M Metadata can be accessed manually (i.e. with human intervention) (Essential)
- RDA-A1-02D Data can be accessed manually (i.e. with human intervention) (Essential)
- RDA- A1-03M Metadata identifier resolves to a metadata record (Essential)
- RDA-A1-03D Data identifier resolves to a digital object (Essential)

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

- RDA-A1-04M Metadata is accessed through standardised protocol (Essential)
- RDA-A1-04D Data is accessible through standardised protocol (Essential)
- RDA-A1-05D Data can be accessed automatically (i.e. by a computer program) (Important)

Principle: "A1.1 the protocol is open, free, and universally implementable".

RDA Indicator:

- RDA-A1.1-01M Metadata is accessible through a free access protocol (Essential)
- RDA-A1.1-01D Data is accessible through a free access protocol (Important)

Principle: *I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.*

FAIRsFAIR Metric

FsF-I1-01M Metadata is represented using a formal knowledge representation language.

This FAIRsFAIR Metric also has a 'B' Level association with R14. Data Re-Use

FsF-I1-02M Metadata uses semantic resources.

This FAIRsFAIR metric also has a 'B' Level association with R14. Data Re-Use

RDA Indicators:

- RDA-I1-01M Metadata uses knowledge representation expressed in standardised format (Important)
- RDA-I1-01D Data uses knowledge representation expressed in standardised format (Important)
- RDA-I1-02M Metadata uses machine-understandable knowledge representation (Important)
- RDA-I1-02D Data uses machine-understandable knowledge representation (Important)

Question: How do you understand and apply machine-actionable knowledge representation to ensure the interoperability of your (meta)data?

Response:

Discussion: Though the Principle addresses Interoperability, the focus of the Principle and the Indicators is on 'standards' in terms of knowledge representation, machine-understandability, self-description etc. This suggests they should either be mapped here to R15 Technical Infrastructure, or that we need to consider these standards as part of higher level context (see Context: Stakeholder Ecosystem & Standards above).

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

***Principle**: I2. (meta)data use vocabularies that follow FAIR principles.*

RDA Indicator:

- RDA-I2-01M Metadata uses FAIR-compliant vocabularies (Important)
- RDA-I2-01D Data uses FAIR-compliant vocabularies (Useful)

***Principle**: "R1.3. (meta)data meet domain-relevant community standards."*

FAIRsFAIR Metric

FsF-R1.3-02D Data is available in a file format recommended by the target research community.

FsF-R1.3-01M Metadata follows a standard recommended by the target research community of the data.

Indicator:

- RDA-R1.3-01M Metadata complies with a community standard (Essential)
- RDA-R1.3-01D Data complies with a community standard (Essential)
- RDA-R1.3-02M Metadata is expressed in compliance with a machine understandable community standard (Essential)
- RDA-R1.3-02D Data is expressed in compliance with a machine-understandable community standard (Important)

Question: what standardised communications protocol do you use to enable retrieval of (meta) data?

Response:

Discussion: Though the Principle addresses Access, the Principle depends on standards, in this case for a communications protocol' while the indicators are a mix of object characteristics and standard requirements. This suggests they should either be mapped here to R15 Technical Infrastructure, or that we need to consider these standards as part of higher level context (see *Context: Stakeholder Ecosystem & Standards* or *Context. Data & Collection Context* above).

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

R16. Security

***Principle**: A1.2 the protocol allows for an authentication and authorization procedure, where necessary*

RDA Indicator

● RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation (Useful)

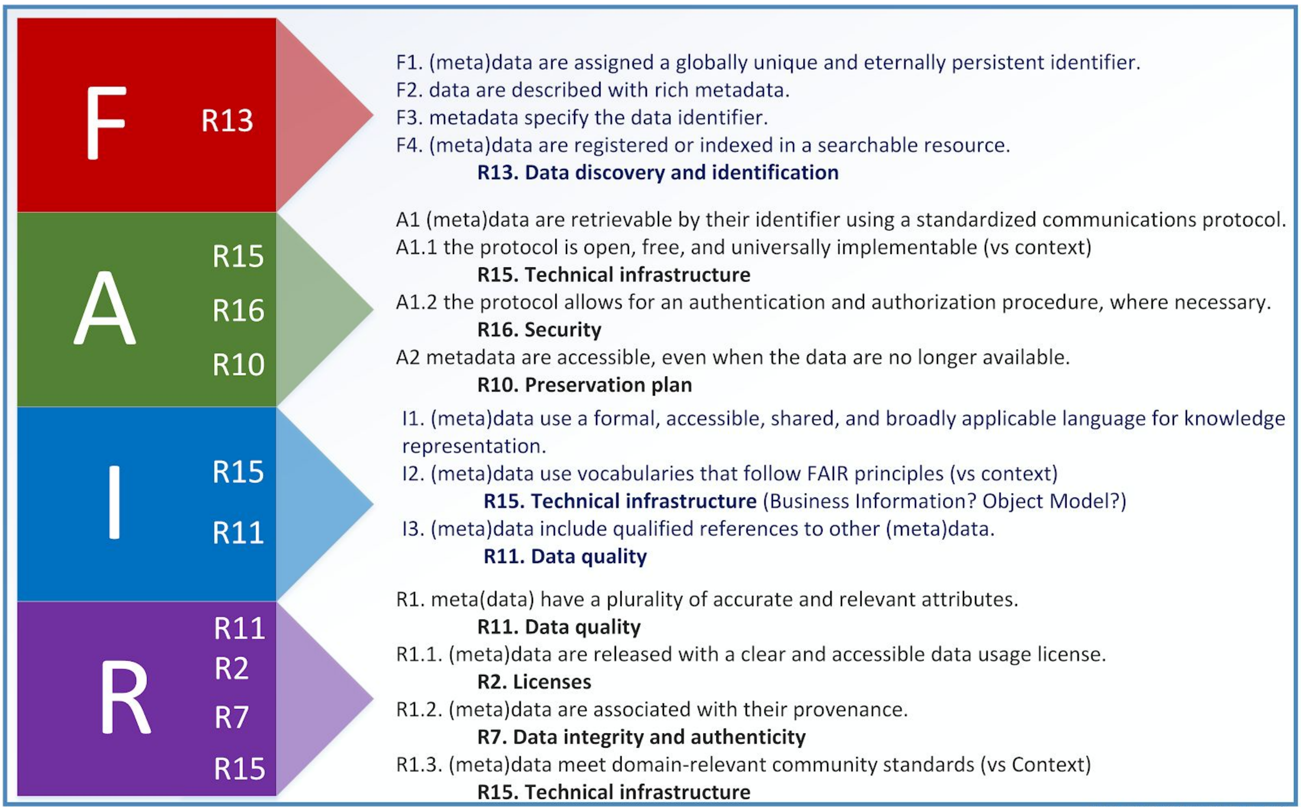
Though the Principle here is Accessibility, the application of authentication/authorisation aligns with R15 Security under CoreTrustSeal.

Question: How do you define the rules for applying authentication and authorisation? Which protocols do you use?

Response:

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

Appendix 1: FAIR Principles to CoreTrustSeal Alignment



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.
R13. Data discovery and identification

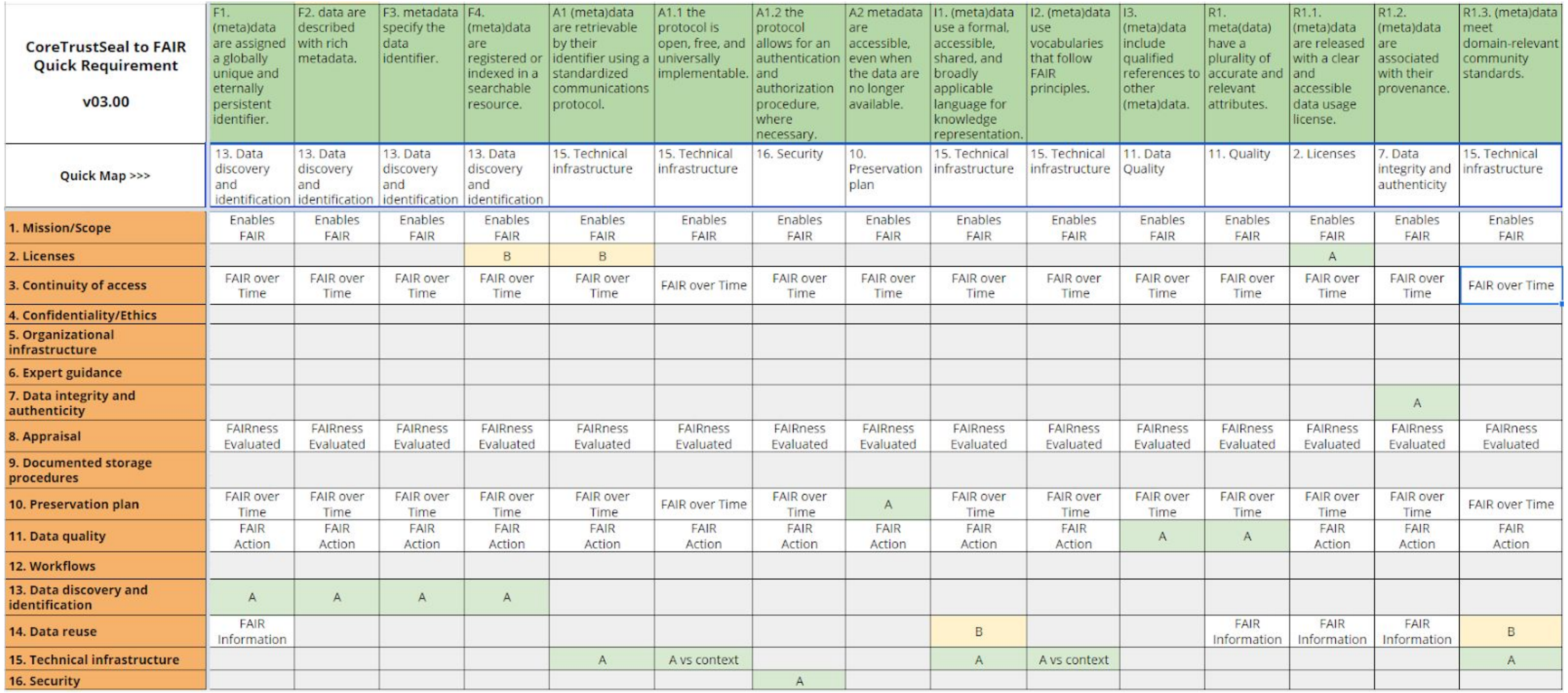
A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable (vs context)
R15. Technical infrastructure
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
R16. Security
A2 metadata are accessible, even when the data are no longer available.
R10. Preservation plan

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles (vs context)
R15. Technical infrastructure (Business Information? Object Model?)
I3. (meta)data include qualified references to other (meta)data.
R11. Data quality

R1. meta(data) have a plurality of accurate and relevant attributes.
R11. Data quality
R1.1. (meta)data are released with a clear and accessible data usage license.
R2. Licenses
R1.2. (meta)data are associated with their provenance.
R7. Data integrity and authenticity
R1.3. (meta)data meet domain-relevant community standards (vs Context)
R15. Technical infrastructure

HLH-CoreTrustSeal-FAIR-Map_02_00.vsdx

Appendix 2: CoreTrustSeal to FAIR Mapping: Quick Reference



CoreTrustSeal to FAIR Quick Requirement **v03.00**	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 communications protocol.	A1.1 the protocol is open, free, and universally implementable.	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-relevant 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				B	B								A		
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														A	
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	A	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	A	A	FAIR Action	FAIR Action	FAIR Action
12. Workflows															
13. Data discovery and identification	A	A	A	A											
14. Data reuse	FAIR Information								B			FAIR Information	FAIR Information	FAIR Information	B
15. Technical infrastructure					A	A vs context			A	A vs context					A
16. Security							A								

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