CoreTrustSeal^{+}FAIR Overview

A FAIRsFAIR Discussion Document

Introduction

This document represents the initial alignment of CoreTrustSeal to FAIR requirements to support repositories seeking to enable FAIR data. The ten repositories receiving support to achieve CoreTrustSeal through the FAIRsFAIR project will respond to these high level questions. Wider feedback is also welcomed. Responses will be integrated into a first draft CoreTrustSeal+FAIR template.

Elaborating CoreTrustSeal R0: Context for FAIR Assessment

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

Additional Context: FAIR Alignment

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 the CoreTrustSeal+FAIR Requirements as they are clarified, contextualised and as indicators are defined and test processes agreed.

This section references each FAIR principle, but does not address the draft indicators[1]. Indicators are currently mapped at the Requirements level.

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:

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

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

Response:

Additional Context: High Level Repository Capabilities

Context: Stakeholder Ecosystem & Standards

Technical standards neatly 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:

Context. Business Information Management: Evidence

A key dependency for any self-assessment and any programme of operational change is a suite 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. How do you integrate data management plans (DMP)?

Response:

Context. Business Systems: Managing Change

The people, processes and technologies which make up data infrastructure must be capable of change over time to meet changes in circumstances and to manage improvement.

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

Response:

Context. Data & Collection Context

A broad understanding of the objects and collections of objects being curated is important to CoreTrustSeal+FAIR.

Question: Briefly describe the characteristics of the data that you curate that you consider important. Describe how the data is divided up into different collections and why (different data, 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 which is not part of the object model?

Response:

CoreTrustSeal Requirements^{+}FAIR

The initial mapping of FAIR Principles to the CoreTrustSeal Requirements is presented in an Appendix.

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
R0. Context. Other Relevant Information

NB: Once agreed the additional items above under "Elaborating CoreTrustSeal R0: Context for FAIR Assessment" will be integrated here.

Organisational Infrastructure

R1. Mission/Scope

Not directly mapped to one or more FAIR Principles

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

Response:

Discussion: As a standard evidence item a mission statement can be mapped to capability levels e.g. 0 (Incomplete-none), 1 (Initial-exists) 2 (managed) 3 (defined as part of the wider organisational processes). 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

Principle 'A*1.2 the protocol allows for an authentication and authorization procedure, where necessary'* depends on criteria set by licences, but A1.2 is mapped to R15 Technical Infrastructure.

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

Indicators:

- R1.1 R1.1-01M Metadata includes information about the licence under which the data can be reused Essential
- R1.1 R1.1-02M Metadata refers to a standard reuse licence Important
- R1.1 R1.1-03M Metadata includes licence information in the appropriate element of the metadata standard used Essential
- R1.1 R1.1-04M Metadata refers to a machine-understandable reuse licence Important
- R1.1 R1.1-05M Metadata includes information about consent for reuse (e.g. for personal data) 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: Rights Management can be mapped to capability levels e.g. 0 (Incomplete-none), 1 (Initial-licences exist) 2 (licences are managed) 3 (licences defined as part of the wider organisational processes). Do all of the indicators defined need to be met to ensure +FAIR? Licences may depend on confidentiality and ethical issues addressed under R4.

Comments:

Note: '*Principle: A1 (meta)data are retrievable by their identifier using a standardized communications protocol'* has an indicator: "A1 A1-01M Metadata includes information about access conditions Essential". In this case the Principle maps to R15 Technical Infrastructure but the indicator aligns with Licences as regards access conditions. This is an example where we need to make practical choices about mapping CoreTrustSeal to FAIR.

Comments:

R3. Continuity of access

Not directly mapped to one or more FAIR Principles

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 focused. *"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.

Discussion: Continuity of Access can be mapped to capability levels e.g. 0 (Incomplete-none), 1 (plan exists) 2 (plan is managed) 3 (plan is integrated as part of the wider organisational processes). 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:

R4. Confidentiality/Ethics

Not directly mapped to one or more FAIR Principles

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: Confidentiality/Ethics can be mapped to capability levels e.g. 0 (Incomplete-none), 1 (aware of issues) 2 (issues managed ad hoc) 3 (plan is integrated as part of the wider organisational processes). Do we need to add anything else here to ensure +FAIR?

Comments:

R5. Organizational infrastructure

Not directly mapped to one or more FAIR Principles

This requirement is primarily repository rather than object focussed.

Discussion: It is not immediately clear how such a potentially complex area as organisation 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

Not directly mapped to one or more FAIR Principles

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: It is possible to map Expert Guidance to capability levels e.g. 0 (Incomplete-no external contact), 1 (some external expertise sought) 2 (consistent community engagement) 3 (organisation-wide plan to seek external expertise) but this is a very broad area in which to apply a single three tier ‘score’. 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:

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*.

Indicators:

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

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

Response:

Discussion: It is possible to map integrity/authenticity to capability levels e.g. 0 (no change control or fixity checks), 1 (basic process in place) 2 (documented version control and change logs) 3 (practice is integrated across the organisation). But should integrity (avoiding unintended change) be addressed separately? Do we need to add anything else here to ensure +FAIR?

Comments:

R8. Appraisal

Not directly mapped to one or more FAIR Principles

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 objects do 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: It is possible to map Appraisal to capability levels e.g. 0 (Incomplete-no set standards), 1 (ad hoc review process) 2 (appraisal and selection rules and associated process) 3 (rules and process integrated into organisation-wide system). How can FAIRness be best evaluated at the point of appraisal/deposit?

Comments:

R9. Documented storage procedures

Not directly mapped to one or more FAIR Principles

The FAIR Principles do not directly address data storage.

Discussion: It is possible to map storage to capability levels e.g. 0 (no bit level assurance), 1 (basic back up mechanisms) 2 (backup and restore measures with N copies in N locations on N media are documented and validated) 3 (integrated and monitored bit level assurance planning across the organisation). Do we need to add anything else here to ensure +FAIR?

Comments:

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*.

> **Indicator**: A2 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: it is possible to map Preservation Planning to capability levels e.g. 0 (Incomplete-no set standards), 1 (preservation risk addressed when detected) 2 (standard procedures for detecting and addressing risk) 3 (rules and process integrated into organisation-wide system covering full collection). What level of additional detail on FAIRness is required to demonstrate preservation of FAIR data characteristics?

Comments:

R11. Data quality

Not directly mapped to one or more FAIR Principles

'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 as part of Stakeholders and Standards under context (above).

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

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

Response:

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

Indicators

- I3 I3-01M Metadata includes references to other metadata Important
- I3 I3-01D Data includes references to other data Important
- I3 I3-02M Metadata includes references to other data Useful
- I3 I3-02D Data includes sufficiently qualified references to other data Useful
- I3 I3-03M Metadata includes sufficiently qualified references to other metadata Important
- I3 I3-04M Metadata include sufficiently 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.

Indicator:

- R1 R1-01M Sufficient metadata is provided to allow reuse, following domain/discipline-specific metadata standard 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? Which, if any, disciplinary or domain standards do you follow?

Response:

Discussion: It is possible to map Quality to capability levels e.g. 0 (Incomplete-no set standards), 1 (ad hoc curation process) 2 (curation rules and associated process) 3 (rules and process integrated into organisation-wide system). How can FAIRness be best evaluated during data quality assurance/curation steps?

Comments:

R12. Workflows

Not directly mapped to one or more FAIR Principles

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: It is possible to map Workflows to capability levels e.g. 0 (Incomplete-no workflows), 1 (some workflows) 2 (comprehensive workflows) 3 (workflows developed and managed through an organisation-wide system). But the management of workflows both 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 at a 'managed' level. 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.

Indicators:

- F1 F1-01M Metadata is identified by a persistent identifier Essential
- F1 F1-01D Data is identified by a persistent identifier Essential
- F1 F1-02M Metadata is identified by a universally unique identifier Essential
- F1 F1-02D Data is identified by a universally 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.*

Indicator:

- F2 F2-01M Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard Important

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.*

Indicator:

- F3 F3-01M Metadata includes the identifier for the data Important

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.*

Indicator:

- F4 F4-01M Metadata is offered/published/exposed 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 addressing these as part of “Context. Data & Collection Context” as 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. Some aspects of FAIR Re-use are addressed more broadly elsewhere under CoreTrustSeal.

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 or as part of higher level context (see Context: Stakeholder Ecosystem & Standards above)

Technology

R15. Technical infrastructure

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

Indicators:

- I1 I1-01M Metadata uses knowledge representation expressed in standardised format Essential
- I1 I1-01D Data uses knowledge representation expressed in standardised format Important

- I1 I1-02M Metadata uses machine-understandable knowledge representation Essential
- I1 I1-02D Data uses machine-understandable knowledge representation Important
- I1 I1-03M Metadata uses self-describing knowledge representation Useful
- I1 I1-03D Data uses self-describing knowledge representation Useful

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:

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

Indicator:

- A1 A1-01M Metadata includes information about access conditions Essential
- A1 A1-01D Data can be accessed manually (i.e. with human intervention) Essential
- A1 A1-02D Data can be accessed automatically (i.e. by a computer program) Important
- A1 A1-02M Metadata identifier resolves to a metadata record Essential
- A1 A1-03D Data identifier resolves to a digital object Essential
- A1 A1-03M Metadata is accessed through standardised protocol Important
- A1 A1-04D Data is accessible through standardised protocol 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:

R16. Security

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

.

Indicators

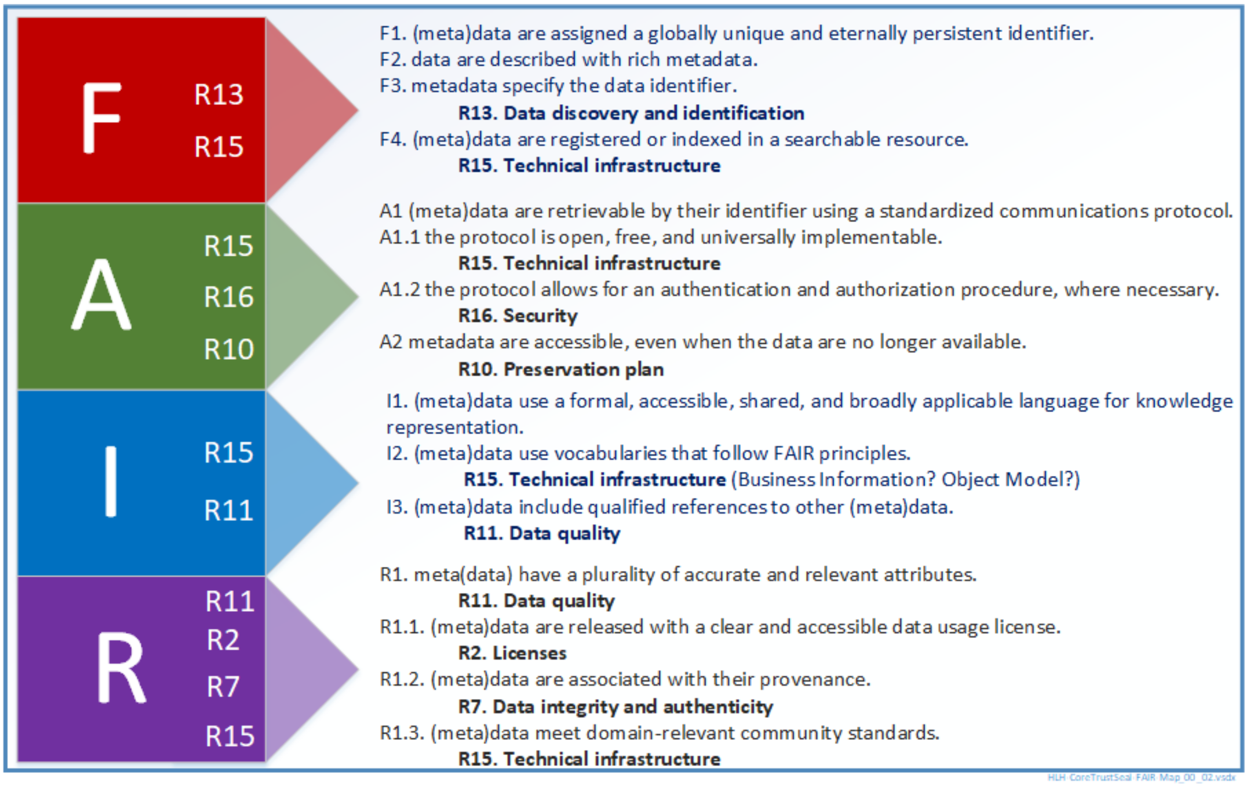
- A1.2 A1.2-01M Metadata includes information relevant for access control Essential
- A1.2 A1.2-01D Data is accessible through an access protocol that supports authentication Useful
- A1.2 A1.2-02D Data is accessible through an access protocol that supports authorisation Useful

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:

Appendix: FAIR 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.
R13. Data discovery and identification
F4. (meta)data are registered or indexed in a searchable resource.
R15. Technical infrastructure

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
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.
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.
R15. Technical infrastructure

HLH CoreTrustSeal FAIR Map_00_02.vsdx