Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity

Work Package	WP4 – FAIR Certification (of Repositories)
Lead Author (Org)	Hervé L'Hours (UKDS)
Contributing Author(s) (Org)	Ilona von Stein, Jerry deVries, Linas Cepinskas (DANS), Robert Huber (UniHB), Joy Davidson, Patricia Herterich (DCC), Benjamin Mathers (UKDS)
Due Date	01.09.2021
Date	31.08.2021
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	10.5281/zenodo.5346822

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)

Versioning and contribution history

Version	Date	Authors	Notes
0.01	2021-07-05	Hervé L'Hours	Initial draft
0.02 to 0.08	2021-08-25	Task 4.1 and Work Package 4 Members	Draft for internal review
1.0	31.08.2021	Task 4.1 and Work Package 4 Members	Content ready

Abstract

This milestone (M4.3) document updates the previous *CoreTrustSeal+FAIR Overview*[1] and the *Draft Maturity Model Based on Extensions and-or Additions to CoreTrustSeal Requirements* (M4.2)[2]. The latter document provides extensive context and references component documents that provide the foundation for this work[3,4,5].

Note: The authors would like to thank the CoreTrustSeal Board for their valuable feedback on a pre-publication version of this text including alignments and target capabilities. The authors acknowledge that while recommending that repositories adopt this approach, no formal adoption and integration into the CoreTrustSeal requirements or processes can take place outside the scheduled, periodic community review process.

Disclaimer

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

Abbreviations and Acronyms

FAIR	Findable, Accessible, Interoperable, Reusable
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
HLAC	High Level Advisory Committee
EGFC	European Group of FAIR Champions
HEIs	Higher Education Institutions
CORDIS	Community Research and Development Information Service

[1] https://doi.org/10.5281/zenodo.4003630
[2] https://doi.org/10.5281/zenodo.4003598
[3] Generic Assessment & Evaluation Reference Model https://doi.org/10.5281/zenodo.3733280
[4] FAIR Ecosystem Components: Vision https://doi.org/10.5281/zenodo.3734273
[5] FAIR Principles: Baseline Comments https://doi.org/10.5281/zenodo.3728131

Table of Contents

1 Overview

Change Log Note: this document covers the fourth iteration (v00.04) of CoreTrustSeal to FAIR mapping (see Appendix for further details) and the initial full release (v00.01) of an associated capability maturity model. Both will be updated to v01.00 based on feedback received.

To ensure that the value of digital assets are maintained it is desirable that trustworthy digital repositories (TDR) enable the deposit, curation and preservation of data that is FAIR for the long term. This milestone document presents updates[6] to the FAIRsFAIR alignment of CoreTrustSeal with repository characteristics that enable FAIR data.

Though many of the CoreTrustSeal Requirements contribute to enabling FAIR data, each FAIR Principle is aligned with a single CoreTrustSeal Requirement to streamline the preparation of self-assessment statements and supporting evidence.

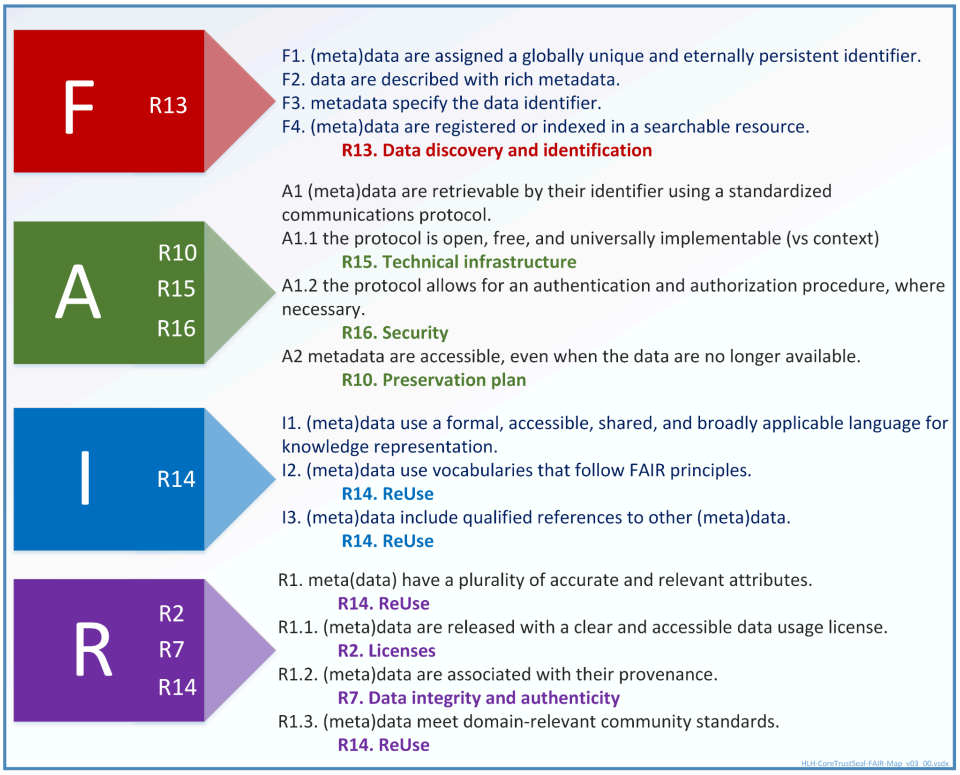

Diagram 1: FAIR to CoreTrustSeal Alignment v00.04

In this text each Requirement to Principle mapping is presented alongside the current iteration of RDA FAIR Data Indicators[7] and the current FAIR tests as implemented by the F-UJI tool[8]. Together these provide all the context necessary for a repository to self-assess as a CoreTrustSeal TDR that enables FAIR data.

Not all repositories will meet the required thresholds for compliance the first time they self-assess. Some repositories will seek to continuously improve beyond the baselines set by the CoreTrustSeal Requirements and FAIR Principles. To support repositories in the identification of their current status

[6] See Appendix: Change Log
[7] FAIR Data Maturity Model. Specification and Guidelines https://doi.org/10.15497/rda00050
[8] [REFERENCE TO D4.5 when released]

and organisational planning towards improvement, CoreTrustSeal+FAIRenabling is accompanied by a capability-maturity (CapMat) assessment approach.

Demonstrating compliance with standards depends on an organisational framework of policies, procedures and other business information[9] in different 'areas of focus'[10]. The first three tiers of the CoreTrustSeal+FAIRenabling CapMat approach are built around the preparation and implementation of appropriate evidence. For an initial self-assessment to a level sufficient to achieve CoreTrustSeal+FAIRenabling the tiers can be understood as follows[11]:

1. Initial: aware of the scope and issue within the area of focus (Requirement/Principle). Lists of all items relevant to the area of focus exist.

2. Managed: processes, procedures and other implementation measures are in place for all items on the lists

3. Defined: managed areas of focus are further integrated into the wider organisational level (policy and practice).

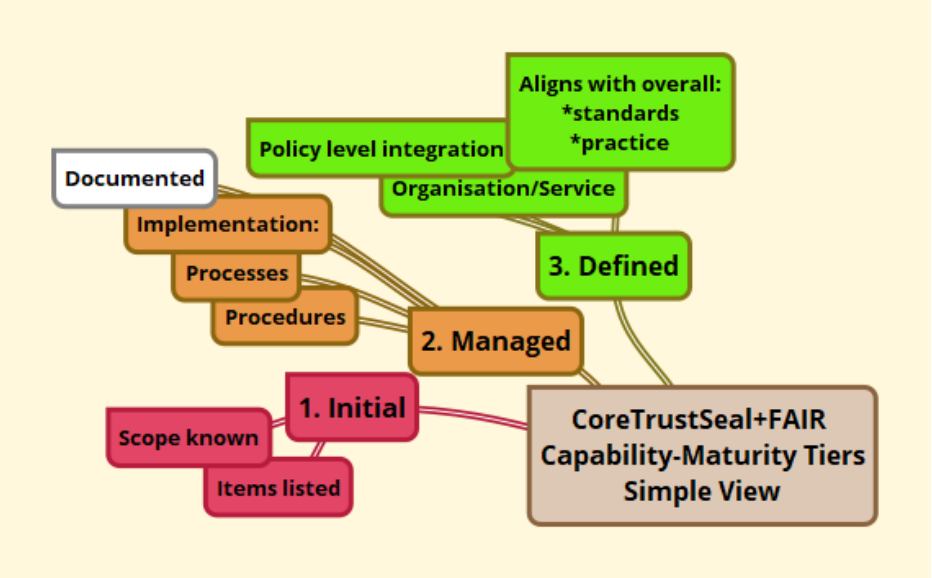

Diagram 2: CoreTrustSeal+FAIR Capability-Maturity: 3 Tier Simple View

The current recommendation is that a capability level of 2. Managed across all Requirements should be sufficient to demonstrate CoreTrustSeal compliance and FAIRenabling. Levels of 3 and above (see diagram below) are highly desirable, but not all repository data services are in a position to influence integration of practice at the policy level.

Once compliance with Requirements/Principles reach *capability* levels of three it becomes more meaningful to talk about overall *maturity* of an organisation or service. Further progress can be made by achieving level 4 (quantitatively managed) or 5 (optimising).

[9] See Appendix: Capability-Maturity & Policy-Evidence Frameworks.
[10] See explanatory diagram in Appendix: Capability-Maturity & Policy-Evidence Frameworks
[11] See Appendix: Capability-Maturity and Community Engagement Descriptions, for formal definitions.

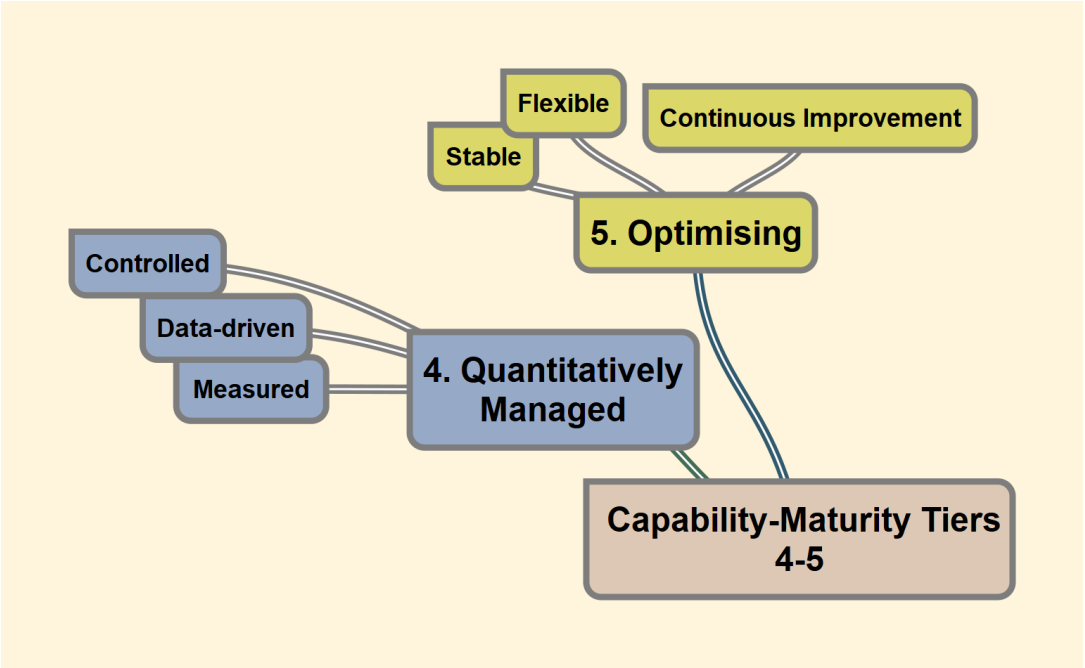

Diagram 3: CoreTrustSeal+FAIR Capability-Maturity. Tiers 4 and 5

Requirement R6. Expert Guidance also proposes evaluation of Community Engagement[12].

Repositories may have extremely heterogeneous collections with some objects falling short of desired criteria for some valid reason e.g. an older but high-value, high demand dataset that it is complex or costly to bring up to standard The human-mediated element of CoreTrustSeal assessment is capable of being more subtle in judgement in such cases compared to an entirely machine-automated process.

Future FAIRsFAIR project deliverables will make proposals for the next revision of the CoreTrustSeal Requirements including the adoption of a +FAIRenabling assessment and the use of the CapMat model. Although discussions around the topic of FAIR data assessment are very active, at present there are no formally managed and endorsed FAIR object assessments or FAIRenabling repository assessment processes. But as noted in the CoreTrustSeal+FAIRenabling Roadmap (Internal draft) "*repositories undertaking self-assessment can find immediate benefit in using the current CoreTrustSeal (v2.0) alongside the CoreTrustSeal+FAIRenabling mappings, the proposed CoreTrustSeal capability/maturity model and FAIR object evaluation tools such as F-UJI*". That document also provides dependencies and a timeline for automated assessment of FAIRenabling characteristics and the integration of tests for digital objects' FAIRness into repository evaluation.

Repositories using CoreTrustSeal+FAIRenabling CapMat should do so in conjunction with the CoreTrustSeal Extended Guidance[13]. Self-assessment and evidence should always demonstrate that the needs of a defined designated community of (meta) data users are being met. Note that CoreTrustSeal always prefers publicly available evidence to support self-assessment statements. Evidence may be public at any level of capability-maturity, but we would suggest that assigning a level 3 status of 'defined' should depend on having appropriately managed public documentation evidence in place.

[12] See Appendix: Capability-Maturity and Community Engagement Descriptions, for formal definitions.
[13] CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022

2 CoreTrustSeal+FAIRenabling CapMat

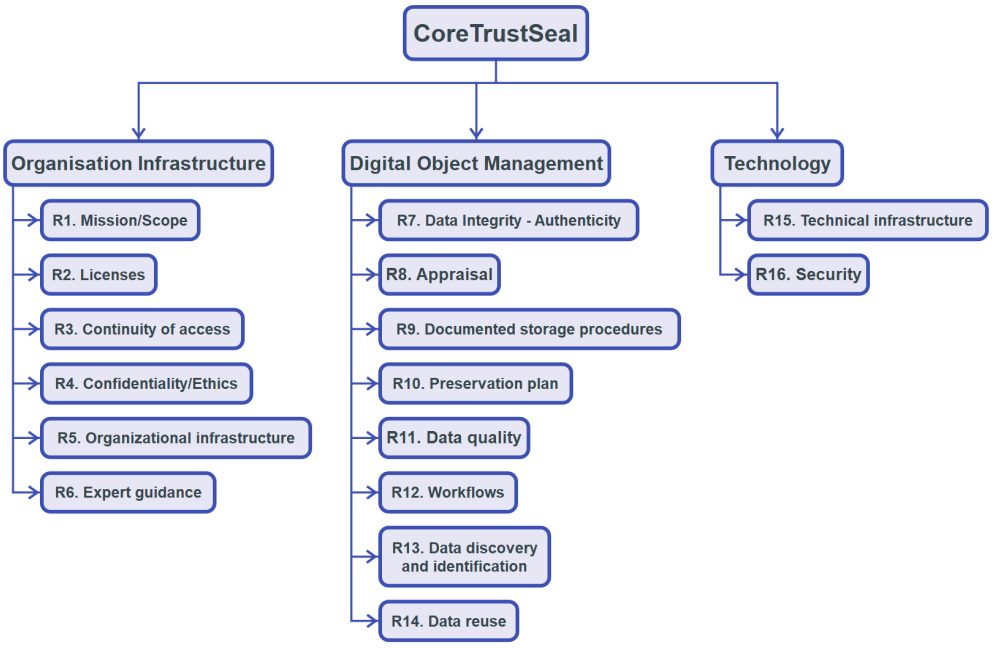

Diagram 4: CoreTrustSeal Requirements

The following sections describe each of the CoreTrustSeal Requirements in terms of capability tiers 1-3 (with additional notes preceded by 'CapMat') and FAIR mappings (with additional comments and guidance preceded by 'FAIRenabling'). Quoted text starting with "Principle:" is adapted from *FAIRsFAIR CoreTrustSeal-plus-FAIR Overview_03_00*[14] including the full Principles text, RDA Indicators and current test descriptions applied by the F-UJI tool.

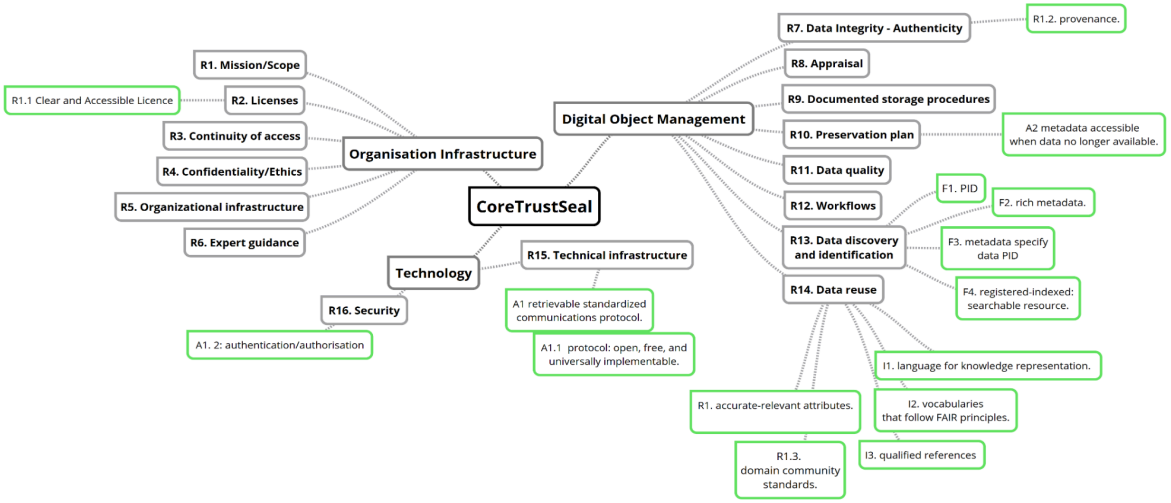

Diagram 5: CoreTrustSeal Requirements to FAIR Principles Alignment (See Appendix 4)

[14] https://doi.org/10.5281/zenodo.4003630

2.1 Organisational Infrastructure

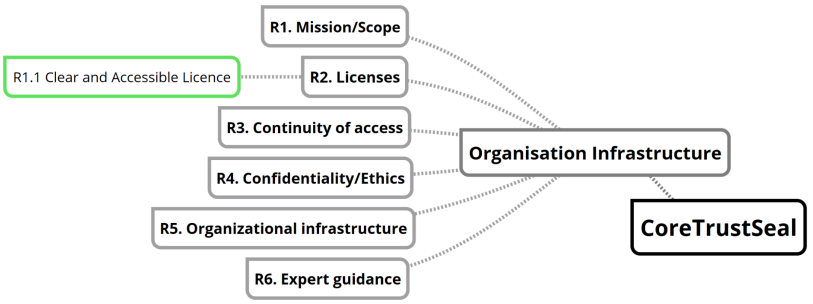

Diagram 6: Organisation Infrastructure to FAIRenabling

2.1.1 R01. Mission/Scope

The key concepts for a data service mission statement might include: designated community, deposit, store, curate, preserve, access and reuse.

1. Initial: Self-assessment statements and evidence reference the key concepts and demonstrate that they are important to the applicant.

2. Managed: A mission statement is in place incorporating locally relevant key concepts.

3. Defined: A formal mission statement exists as part of a policy framework that ensures it is aligned across repository practice. It is clear who approves the mission statement and how it is reviewed and revised over time.

CapMat: For a mission statement to be *quantitatively managed* (level 4) it would need to be linked to functions and activities that ensure it is delivered to defined levels (e.g. KPI). These would permit *optimisation* (level 5)

2.1.2 R02 Licenses

1. Initial: Every digital object being curated is known and recorded. The repository is aware that a rights statement is required for each, potentially with different rights applying to different parts of the data and metadata.

2. Managed: Every digital object, and relevant part of a digital object, has clear rights associated with it; defining permissions, prohibitions and duties of depositors, repository and end users. These rights permit the repository to store, curate, preserve and provide access to the digital objects for the defined period of responsibility (including 'indefinite') whether or not the original depositor or other rights holders remain available.

3. Defined: Rights management is integrated with the wider policy and practice framework and is managed in line with internal and external changes that impact rights.

CapMat: Reaching level 4 or 5 would be more practical if licences were described using a structured or machine-actionable standard (e.g. ODRL).

+FAIRenabling: digital object metadata includes license information covering (meta)data reuse.

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

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)"

2.1.3 R03. Continuity of Access

1. Initial: Every digital object being curated, and every function that delivers the service around those objects is understood. The level of curation and level of service required to maintain these over time is known.

2. Managed: Policies and procedures are in place for each function. These go beyond the day to day process definitions (R12). For R03 they consider how the impact of a disaster can be mitigated, minimised and recovered from. This level would permit a succession plan to be *developed* for handover of digital objects and related functions in the event that the repository ceased to function.

3. Defined: Business continuity and disaster recovery measures are integrated across the whole organisation. This may be a necessary level of capability for the successful *implementation* of a succession plan.

CapMat: because a complete succession plan covering all repository functions that is formally agreed with a successor organisation[15] is a high bar for many repositories; this has been included in level 3, though it may be possible at level 2. Risk to digital objects and related functions are of course minimised with continuous improvement (5. Optimised).

2.1.4 R04. Confidentiality/Ethics

1. Initial: Repositories know all the relevant legislation and ethical policies and practices that apply to the functions they offer and their digital objects.

2. Managed: The characteristics of each digital object (e.g. contains sensitive data, access restricted to a geographic area) are associated with relevant legal and ethical standards. Functions applied to those objects meet those legal and ethical standards.

3. Defined: Legal and ethical practice is integrated into a whole-organisation policy and procedural framework.

[15]Necessary to reach the CoreTrustSeal Compliance Level 4. 'fully implemented'

2.1.5 R05. Organisational Infrastructure

1. Initial: The applicant is clear on what *they* are responsible for, when responsibility is with a host organisation (if present) and when responsibility is shared with a third party (NB: this is the minimum necessary for a clear 'insource/outsource' statement in R0. Context). Staff names, job titles and role descriptions are in place. Departmental names and function descriptions are in place. Projects and groups are listed and their purposes and intended outcomes are known. Individual, departmental, project and group costs and budgets are known.

2. Managed: Any hierarchies and decision making workflows are documented. Organisational structure descriptions or diagrams exist. Human and financial resources are managed in line with relevant workloads and funding availability. Funding is sustainable and sufficient.

3. Defined: Governance and resources are managed across organisational policy and practice. Cross-sectional alignment is in place across repository data curation and preservation, governance, technology and security.

CapMat: Quantitatively managed would equate to monitoring of both time and costs against repository functions with metrics (e.g. KPI) in place. Optimising (5) would involve the continuous improvement of governance processes and resource management to maximise service levels and minimise costs.

2.1.6 R06. Expert Guidance

This Requirement contains two areas for evaluation: 1. That the expertise needed is understood and is either delivered internally (R05) or sought externally (R06) and 2. That the repository engages with the wider community in relevant areas of practice.

Internal and External Expertise

1. Initial: The repository has listed the knowledge, skills and expertise related to their (meta) data and functions. This includes technical infrastructure to the degree implied by the scope of R15.

2. Managed: The organisation monitors and maintains processes and procedures, ensuring the appropriate internal knowledge skills and expertise remain available to deliver them. Defined relationships with external service or information providers are in place to provide relevant guidance and expertise.

3. Defined: Alignment of objects and functions with internal and external expertise is integrated into the wider organisational policy and practice. These include managed feedback from the designated community that is explicitly met with responses that address their evolving needs.

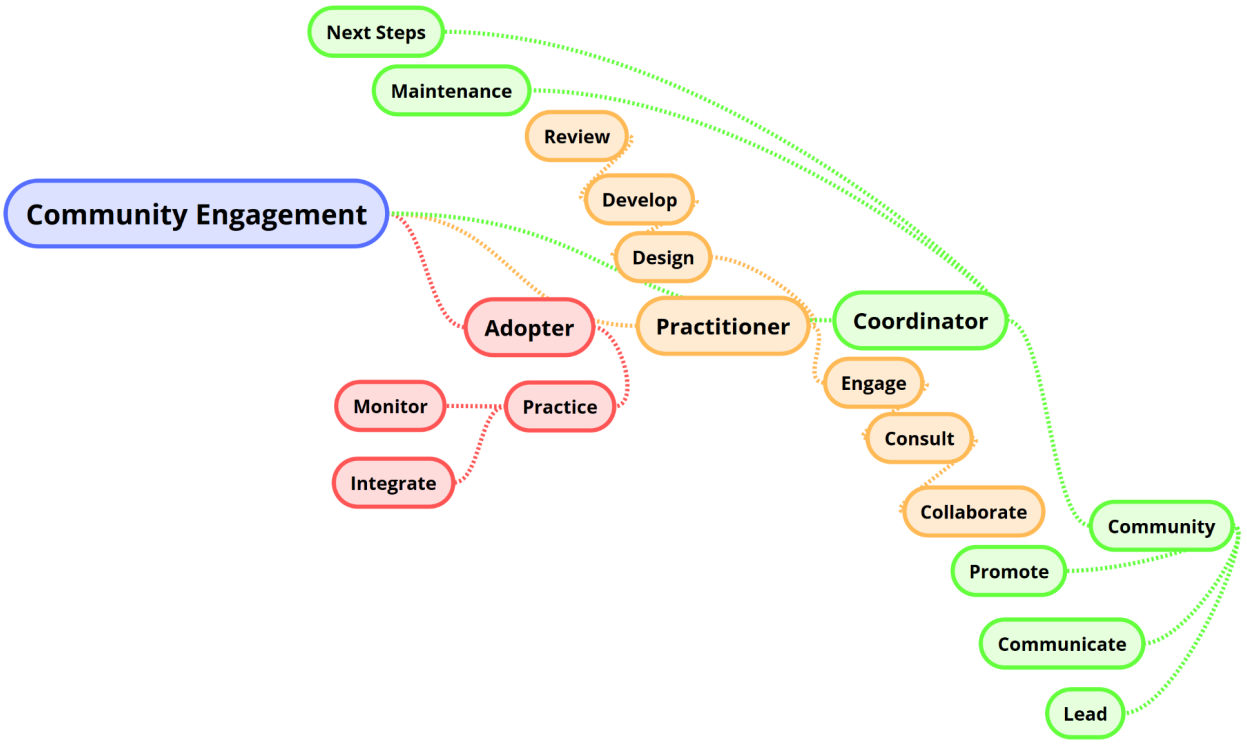

Diagram 7: Community Engagement Tiers

Adopter: For each area of expertise, define how the repository monitors community practice and integrates it into local practice.

Practitioner: In addition to adoption, the repository also engages with the design, development, and review of community practice. Consults and collaborates widely.

Coordinator: The repository is an adopter and practitioner that also takes a community coordination and leadership role. Driving maintenance and updates of existing practice and identifying next steps for the development of policy and implementation standards. Actively communicating and promoting existing and emerging approaches to the immediately impacted communities and the wider data infrastructure landscape.

Community Engagement: 'Adopter' should be the minimum target for a CoreTrustSeal applicant. Increased community engagement increases confidence in the repository's overall CoreTrustSeal+FAIR application. These tiers have been developed from the basis of the three community engagement tiers provided in *Appendix: Capability-Maturity and Community Engagement Descriptions*.

2.2 Digital Object Management

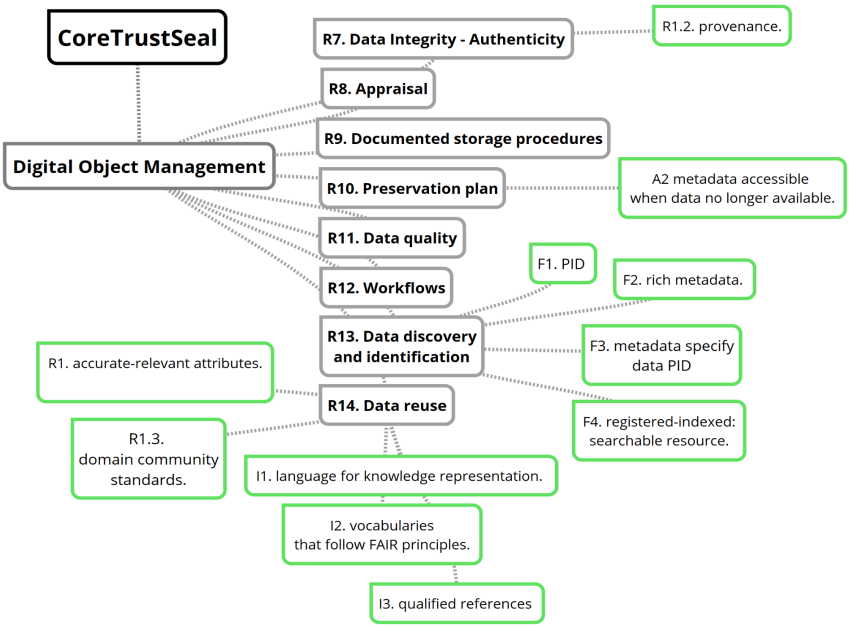

Diagram 8: Digital Object Management to FAIRenabling

2.2.1 R07. Integrity and Authenticity

Integrity

1. Initial: Objects are subject to integrity checks at the point of deposit, transfer to archival storage and transfer to access. Stored objects are subject to periodic integrity checks.

2. Managed: Integrity measures are aligned with processes and procedures. Any functions where change is not specifically permitted are supported by assurance that unintended change is avoided.

3. Defined: Integrity measures are part of an overall policy and procedural framework that defines which actors and agents are responsible for ensuring integrity and this is assured through governance and technical measures.

Authenticity

1. Initial: The minimum level of pre-repository digital object provenance that is required by the designated community is available (cf: R08 Appraisal). Permitted and required changes to the object made during repository custody are listed.

2. Managed: Changes made to objects, whether originals or copies, follow documented processes, and changes are recorded. A documented version system is in place. Relevant information is made available to end users. This depends on a rights framework being in place (R2) and actors and their roles being known (R04).

3. Defined: All changes are part of an overall policy and procedural framework that defines which actors and agents can make them, enforces this through governance and technical measures and records actions and outcomes at a granular level.

CapMat: Though presented together in CoreTrustSeal 2.0 the concepts of integrity and authenticity are sufficiently distinct to require separate capability-maturity tiers.

+FAIRenabling: Applicant confirms that metadata includes provenance information about data creation and curation in line with the needs of the designated community.

***"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)"

2.2.2 R08. Appraisal

1. Initial: The minimal, acceptable and ideal characteristics of digital objects to be accepted into the repository are known.

2. Managed: A selection and appraisal process is in place that checks each digital object that is considered for deposit. Preferred and acceptable file format lists and metadata standards/schemas exist. Minimum metadata and documentation at the point of deposit are defined. Any objects that fail to meet the acceptance criteria are rejected, or the reasons for exceptions are documented.

3. Defined: Appraisal and selection processes at the point of deposit are integrated into wider organisational policy and practice. Pre-Repository data management plans (DMP) are integrated at the point of deposit and appraisal outcomes feed into curation and preservation plans.

2.2.3 R09. Storage

1. Initial: Storage locations and media for all digital objects are known through deposit, curation, archival storage, discovery and access (and re-use if mediated by the repository). All storage locations form part of a backup system with at least two copies in separate locations. Backup frequency is known for each location.

2. Managed: All storage media management follows processes, procedures and other implementation measures including management as part of overall information technology infrastructure and technical watch (R15). Storage media are monitored for capacity and failure and are subject to a periodic media refreshment plan. Storage media types are assessed for obsolescence.

3. Defined: Storage is integrated into an overall technical infrastructure management plan which is in turn integrated into the wider organisational policy and practice. Storage locations are assessed for disaster threats. Storage media types, location risks, numbers of copies and backup periodicity all meet a documented minimum threshold (e.g. NDSA levels of Preservation[16] for Storage).

2.2.4 R09. Preservation Plan

1. Initial: Preferred and accepted deposit (R08), access (R14) and preservation file formats and metadata standards/schema are listed. Every (meta) data object in the repository is listed and their file format and metadata standards/schemas are known. Any minimum periods of retention (bit level assurance) are documented.

2. Managed: The curation and preservation levels of all digital objects are documented. The needs of the designated community and the wider technical dependencies (risks) for (meta) data are monitored and used to define and implement curation and preservation actions. Actions, including normalisation, migration, emulation and updates to metadata and documentation ensure (meta) data remain findable, accessible, interoperable and reusable in line with the needs of the designated community. Preservation actions are taken as soon as is practical, in response to or in advance of identified changes to circumstances. Any minimum periods of preservation are documented.

3. Defined: Preservation planning is integrated into the wider organisational policy and practice including governance, resourcing, expert guidance (with community engagement) and technical infrastructure.

CapMat: Preservation planning is a wide topic that has dependencies in many other CoreTrustSeal Requirements. For this reason the proposed capability-maturity tiers focus on preservation actions. The fact that preservation is the central focus of repository data services would suggest that 'defined' should be the minimum level necessary for CoreTrustSeal, though some repositories, particularly those that are hosted by larger organisations may find it hard to deliver preservation planning that is fully integrated with their wider organisational practice.

Ideally preservation planning would be both quantitatively managed (prioritisation based on quantified demand from the community and quantified risk from technology watch) and optimised (continuous improvement and agile responses to opportunity and the need for change). But this ideal has some dependencies on wider community agreement on minimum standards and is likely to require targeted investment in repositories as part of research infrastructure uplift.

+FAIRenabling: The repository preservation policy ensures that metadata is preserved even when an object is removed from your repository? Any exceptions are defined and documented.

*"**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

[16] https://ndsa.org/publications/levels-of-digital-preservation/

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

CoreTrustSeal+FAIR and CapMat note: Principle A2 and its indicator are an explicit requirement that metadata is *preserved* after data is unavailable so it is mapped to R10. But this is also associated with standard practice for persistent identifier management (R13 Data Discovery and Identification). The 'tombstoning' of metadata records is included in the CapMat tiers for R13.

2.2.5 R11. Quality

1. Initial: The repository is aware of the expectations at the point of reuse (R14). Curation activities ensure that any quality levels not met at the point of deposit are addressed to meet reuse and preservation needs. The repository is aware of relevant standards and works to meet them.

2. Managed: The quality expectations of digital objects not met at the point of deposit are integrated into a curation plan. Curation actions take place against the defined standards. Quality assurance of standards compliance takes place and any exceptions are documented. The digital objects at the point of access either reach defined quality thresholds or reasons for not meeting quality standards are documented.

3. Defined: Standards selection and quality assurance measures are integrated into the wider organisational policy and practice.

CapMat: Quality is a challenging concept to define and apply without further context e.g. Bruce & Hillman (2004[17]). Quality and quality assurance depend on the selection or development of appropriate standards (implied across CoreTrustSeal) and workflows (R12) to curate for and check for expected levels of quality.

2.2.6 R12. Workflows

1. Initial: The repository is aware of the processes in place for deposit, curation, preservation, archival storage, discovery and access.

2. Managed: Documented workflows exist for deposit, curation, preservation, archival storage, discovery and access. Curation actions take place in line with defined standards and a curation plan developed at the point of deposit.

3. Defined: Workflow design, management and change management are integrated into the wider organisational policy and practice.

2.2.7 R13. Discovery & Identification

Note: Though presented together in CoreTrustSeal 2.0 the concepts of discovery and identification are sufficiently distinct to require separate capability-maturity tiers and separate alignment with the FAIR Principles.

[17] The Continuum of Metadata Quality: Defining, Expressing, Exploiting

Discovery

1. Initial: The digital objects' data, metadata and documentation are structured and presented in a way that passively permits indexing and harvesting by resource discovery systems.

2. Managed: Standards compliance processes are in place to ensure that (meta)data can be included in resource discovery systems suitable for the designated community. These may include local systems, 'pushing' metadata to third party systems and data catalogues or providing standardised metadata that can be pulled or harvested by other systems (e.g. OAI-PMH). Processes, procedures and other implementation measures are in place for all items on the lists

3. Defined: Managed areas of focus are further integrated into the wider organisational policy and practice.

+FAIRenabling: The repository provides evidence that resource discovery metadata is sufficient for their designated community of users. A disciplinary repository may be expected to provide information for both general purpose resource discovery systems (exposure for indexing by search engines, high level metadata such as Dublin Core or DataCite), and metadata to support their more specialist designated community of users.

*"**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)"

+FAIRenabling: This mapping focuses on the Principle.

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

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)"

Identification

1. Initial: Every (meta)data object in the repository is known and has its own identifier in place.

2. Managed: Every object identifier is locally unique and persists over time. Processes are in place to ensure that the identifier continues to resolve correctly over time and that a metadata record persists even if, for some reason, the digital object is no longer accessible.

3. Defined: Local practice aligns with community agreed minimal standards for designing and managing globally unique identifiers and resolution systems including minimal practice for ensuring persistence and handling 'tombstone' metadata records for objects that are no longer accessible.

+FAIRenabling: All objects in the repository are persistently identified. Any exceptions are documented and explained, including a timetable for complete coverage of persistent identifiers.

*"**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.

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)"

+FAIRenabling: The repository provides evidence that digital object metadata includes its persistent identifier.

***Principle**: F3. metadata specifies 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)

2.2.8 R14. ReUse

The ReUse Requirement of CoreTrustSeal assesses two broad areas:

1. The means by which the repository identifies and addresses the needs of the designated community.

2. The characteristics of the data and metadata in the digital objects that meet these needs.

1. Initial: The repository has a definition of their designated community and documents assumptions about that community's specific needs. The repository documents the formats, metadata standards and other requirements for representing (viewing, editing etc) the (meta) data and is aware that these may need to be updated over time.

2. Managed: The repository actively engages with their designated community to identify their needs in terms of (meta)data usability. This 'community watch' activity is applied alongside a 'technology watch' function that monitors potential risks to the current (meta)data approaches used for digital objects including obsolescence and seeks equivalent or improved alternatives.

3. Defined: Monitoring and change related to continued reusability of digital objects is integrated into the wider organisational policy and practice.

+FAIRenabling: The repository describes how digital objects are linked to other data and metadata to meet the needs of the designated community.

*"**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)

+FAIRenabling: The repository provides evidence that the data and metadata are provided according to the standards of their designated community.

*"**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)"

+FAIRenabling: The repository describes how the metadata provided for digital objects meets the needs of their designated community.

*"**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)"

+FAIRenabling: The repository describes the knowledge representation approaches (schemas, ontologies etc) they use to ensure machine-actionable interoperability.

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.
FsF-I1-02M	Metadata uses semantic resources.

RDA Indicators:

- RDA-I1-01M — Metadata uses knowledge representation expressed in standardized 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)

+FAIRenabling: (as for I1) the repository describes the knowledge representation approaches (schemas, ontologies etc) they use to ensure machine-actionable interoperability.

***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.

RDA 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)

2.3 Technology

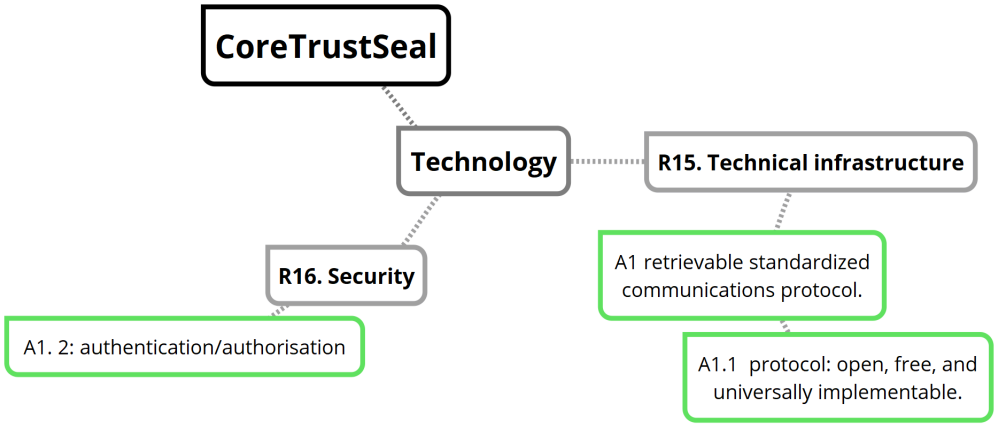

Diagram 9: Technology to FAIRenabling

2.3.1 R15. Technical Infrastructure

1. Initial: Parts of technical systems are listed, issues with technical systems and responses to issues are listed.

2. Managed: Lists of components and services, lists of issues (bug fixes, change requests) and research and development projects are managed through processes, procedures and other implementation measures.

3. Defined: IT service management is integrated with overall repository, governance and resource management.

+FAIRenabling: The repository describes the method by which objects can be accessed.

*"**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.

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)
- 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)"

+FAIRenabling: (as for A1) the repository describes the method by which objects can be accessed.

"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)"

2.3.2 R16. Security

Security is a similarly broad area to technical infrastructure, and as noted above expectations would increase for repositories curating sensitive data. In a similar approach to R15, the reverse engineering of minimal and ideal practice from more advanced security standards (e.g. ISO27001) may be the best approach to defining the 'core' and from there a set of capability-maturity tiers.

1. Initial: The scope of security is defined, any security issues and responses to issues are listed.

2. Managed: Potential threats to the repository and its data, products, services, and users are analysed and risks assessed. Processes are in place to periodically review risk, to mitigate risks where possible and to minimise and respond to threats (whether malicious or through human error) to the IT infrastructure.

3. Defined: Information security management is integrated with overall repository, governance and resource management.

+FAIRenabling: The repository defines their terms for applying authentication and authorisation and the protocol in place to apply access control.

*"**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)"

Appendix 1: Change Log-CoreTrustSeal to FAIR Alignment

Change Log Note: this document covers the fourth iteration (v00.04) of CoreTrustSeal to FAIR mapping. Previous versions were released as CoreTrustSeal plus FAIR Overview[18]. The most recent changes are described below. Feedback to this document will result in a v01.00 release with additional versions released as necessary during the project timescale.

This iteration of the CoreTrustSeal to FAIR alignment has benefitted from engaged feedback from members of the CoreTrustSeal Board. Though many of the CoreTrustSeal Requirements contribute to enabling FAIR data, so there are multiple possible alignments, a single mapping (One FAIR Principle to One CoreTrustSeal Requirement) has been identified to simplify integrating statements and evidence about FAIR enabling into the CoreTrustSeal self-assessment process.

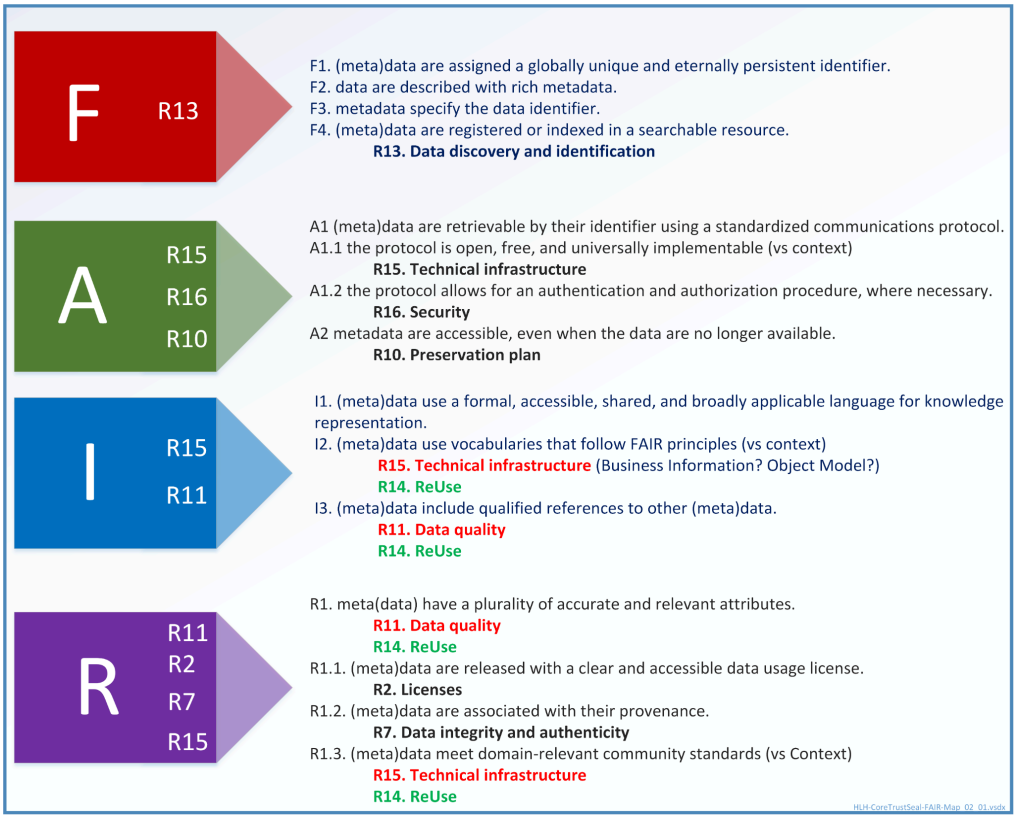

Diagram 10: ***FAIR to CoreTrustSeal.*** Prior mappings in red, new proposed mappings in green

The diagram above is based on the version used in the M4.2 Draft Maturity Model Based on Extensions and-or Additions to CoreTrustSeal Requirements[19] M4.2 2020-09 and *FAIRsFAIR-CoreTrustSeal-plus-FAIR Overview_03_00*[20] with updates to reflect the most recent revised mappings.

Amendments in this version provide a stronger focus on data reuse as a target outcome for long-term FAIR data enabling.

[18] https://doi.org/10.5281/zenodo.373489
[19] https://doi.org/10.5281/zenodo.4003598
[20] https://doi.org/10.5281/zenodo.4003630

Principles I1 and I2 (previously R15 Technical Infrastructure) and Principle I3 (Previously R11. Data Quality) are moved to CoreTrustSeal R14: ReUse.

- Principle: I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- Principle: I2. (meta)data use vocabularies that follow FAIR principles.
- Principle: I3. (meta)data include qualified references to other (meta)data.

Principles R1 (Previously R11 Data Quality) and R1.3 (Previously R15 Technical Infrastructure) are moved to CoreTrustSeal R14: ReUse

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

Principle: “R1.3. (meta)data meet domain-relevant community standards.”

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.” The previous mapping to ‘Quality’ was because this is where the curation work was undertaken. The decision has been taken to focus on digital object information, community information and associated standards (including disciplinary formats and metadata where relevant) under R14 as this provides the best alignment when looking for evidence of fitness of data for ReUse. These amendments also clarify the focus of R15 the ‘IT Service Management’ aspects of standards.

Appendix 2: Capability-Maturity & Policy-Evidence Frameworks

Another text (internal working draft) is looking at proposing an approach for organisations or services designing policy and evidence frameworks. Policy/procedure 'artefacts' (another type of object) define activities and specify approaches and outcomes. They may also provide evidence for compliance with targets (e.g. key performance indicators), standards, or assessments, such as CoreTrustSeal.

In an ideal situation a collection of functions (a service or whole organisation) benefits from defined goals and operational parameters (policies) that are implemented using standard operating procedures to deliver measurable outcomes. In the real world different parts of an organisation or service evolve at different rates. A particular area of focus (e.g. a Requirement from CoreTrustSeal or a Principle from FAIR) may be **managed** in isolation (e.g. due to being prioritised) but only become **defined** in terms of the wider service or organisational practice at a later stage. Demonstrating a higher level of **capability** across a number of areas of focus (moving from *managed* to *defined*) can help to define the overall **maturity** of the service or the organisation.

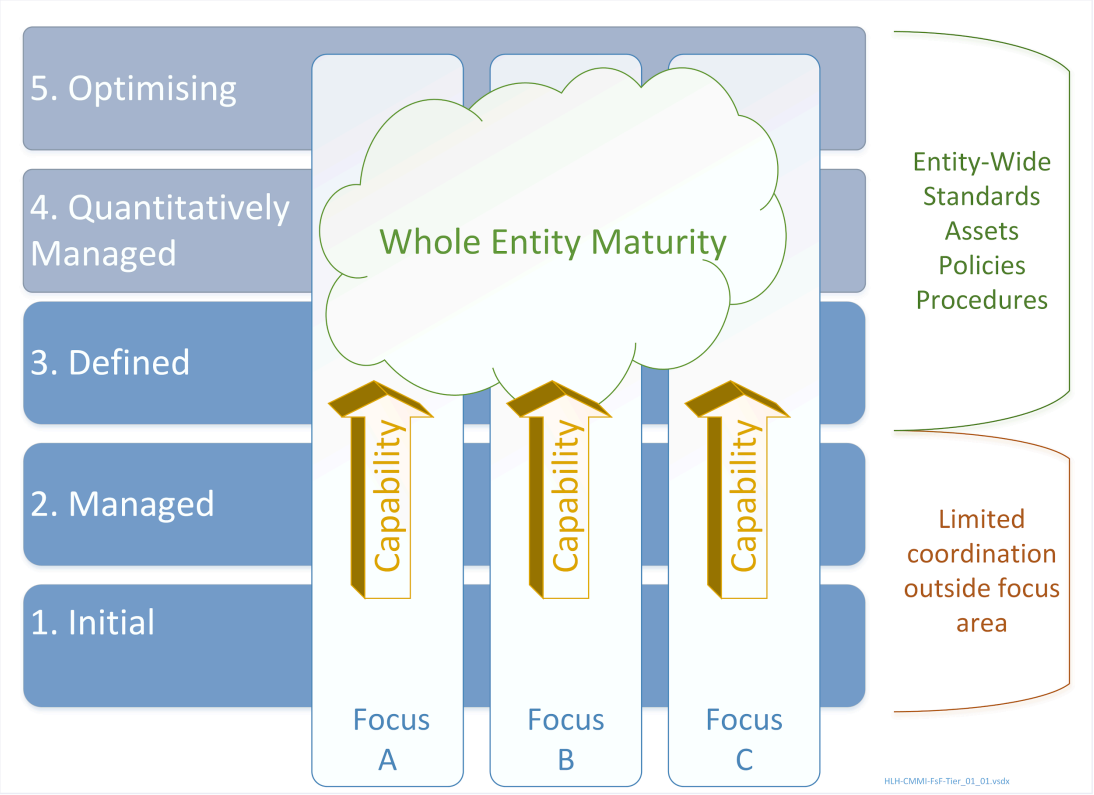

Diagram 11: Comparison of focus areas vs entity-wide tiers

Some capability-maturity approaches include a very low initial level. CMMI uses a 'level 0' for: 'incomplete, ad hoc and unknown, work may or may not get complete'.

For CoreTrustSeal+FAIR we will focus on the tiers and recommend these for self-assessments against the CoreTrustSeal Requirements. Formal definitions are provided in the appendix but for an initial self-assessment they might be considered as follows:

1. Initial: Aware of the scope and issue within the area of focus. Lists of all items relevant to the area of focus exist.

2. Managed: Processes, procedures and other implementation measures are in place for all items on the lists

3. Defined: Managed areas of focus are further integrated into the wider organisational policy and practice.

Beyond these tiers, an area of focus may be *4. quantitatively managed* by being measured and controlled for quality and performance. The data-driven nature of level 4 is a necessary foundation for *5. Optimising* where more flexible responsiveness to change, continuous improvement and innovation are delivered. In some cases we provide additional notes on tiers 4 and 5 to support more advanced assessment, prioritisation and development[21] within organisations and services.

FAIRsFAIR work to date indicates that achieving levels 2 or 3 are sufficient for most CoreTrustSeal Requirements though higher levels are clearly desirable as they will improve organisational performance and reduce the risks to (meta)data.

[21] Levels 4 and above may be dependencies for efficient interoperability between repositories and other data related services, including other repositories. This level of organisational or service interoperability is highly relevant to consolidated research infrastructure design and management (e.g EOSC) but is out of scope for the CoreTrustSeal unless as an expressed need from the designated community.

Appendix 3: Capability-Maturity and Community Engagement Descriptions

A separate text provides a design statement[22] for the FAIRsFAIR approach to tiers of capability and maturity. This was developed to provide internal consistency in project work, in cooperation with discussions around the proposed Science Europe maturity matrices[23] and with a view to alignment with formal capability maturity models such as CMMI[24].

FAIRsFAIR CoreTrustSeal+FAIR Capability-Maturity Definitions

Unless otherwise referenced quoted text is taken from the *FAIRsFAIR Capability/Maturity and Community Engagement Design Statement*[25].

"Compatible but simplified FAIRsFAIR approach (based on the CMMI levels below). Each tier description is applied to the organisation, repository or service entity being evaluated:

1. **Initial**. May be incomplete and falling short of the intent of the area of focus. Aware of and addressing performance issues.

2. **Managed**. Limited but complete coverage that delivers the full intent of the area of focus. Although lacking full alignment with overall organisational standards and practice, Identifies and monitors performance objectives. Includes and builds on level 1.

3. **Defined**. Complete coverage that delivers the full intent of the area of focus and aligns with overall organisational standards and practice. Identifies and monitors performance objectives that expand alignment to the whole organisation.

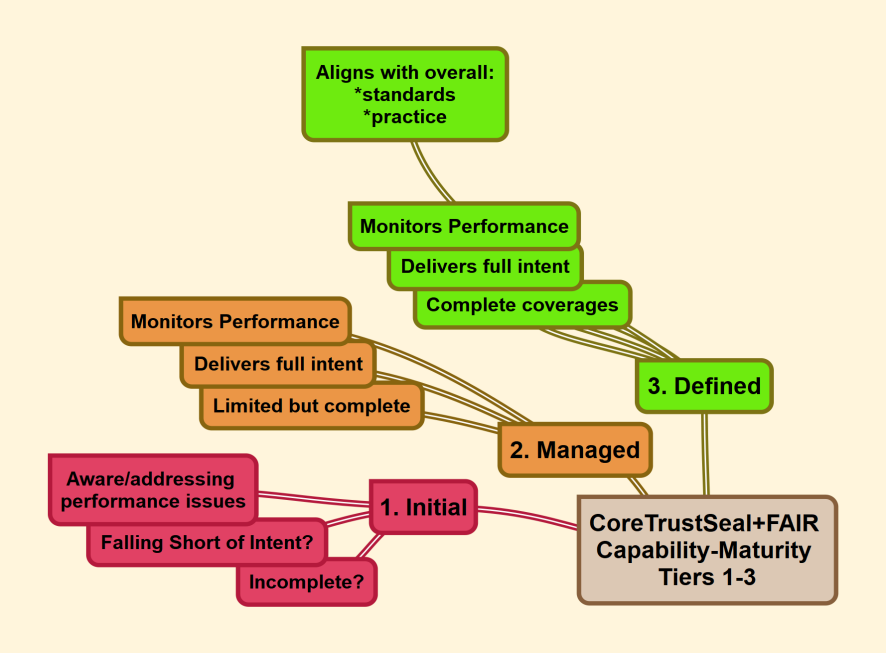

Diagram 12: CoreTrustSeal+FAIR Capability-Maturity: 3 Tier View

[22] https://doi.org/10.5281/zenodo.4705235

[23] https://doi.org/10.5281/zenodo.4769702
[24] Capability Maturity Model Integration (CMMI)
[25] https://doi.org/10.5281/zenodo.4705235

The following definitions are taken from CMMI 2.0[26].

"**Level 4: Quantitatively Managed**. Measured and controlled. Organization is data-driven with quantitative performance improvement objectives that are predictable and align to meet the needs of internal and external stakeholders.

Level 5: Optimizing. Stable and flexible. Organization is focused on continuous improvement and is built to pivot and respond to opportunity and change. The organization's stability provides a platform for agility and innovation."

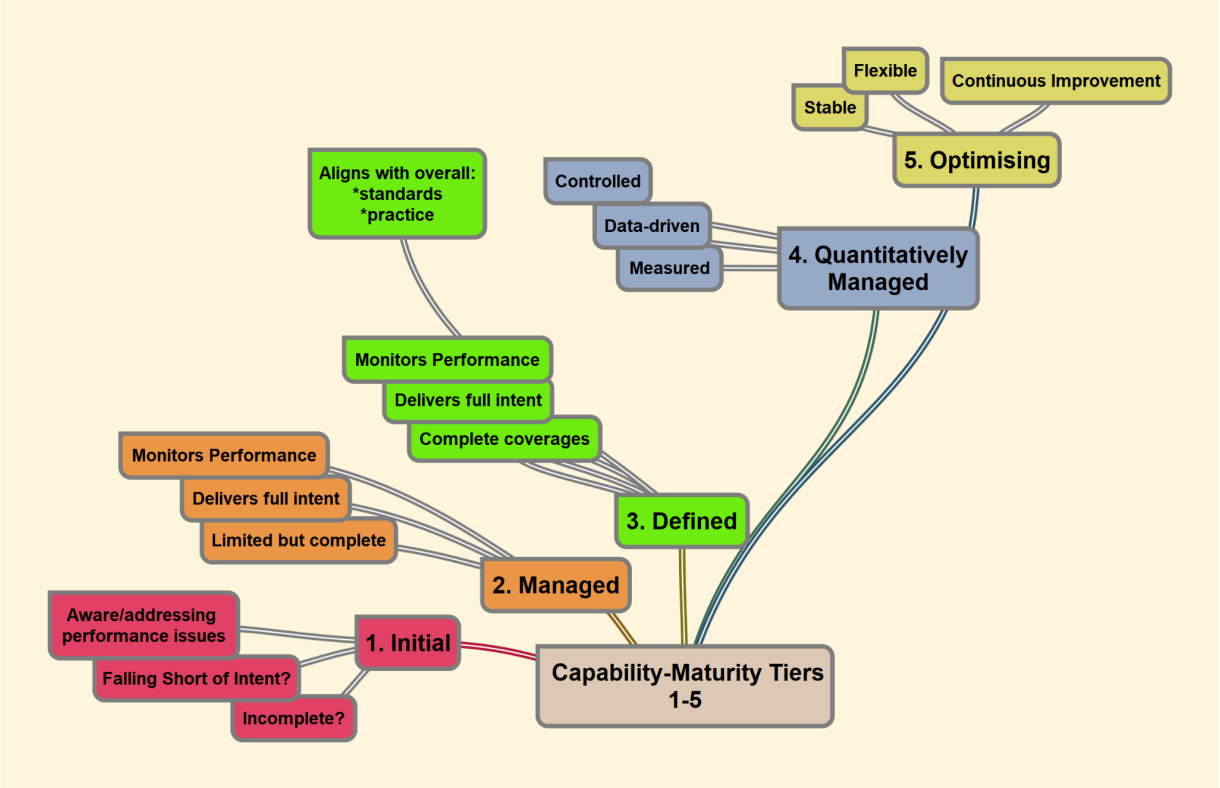

Diagram 13: Tiers 1-5

Community Engagement Definitions:

"Awareness: Monitors community practice and makes local practitioners aware of it.

Adoption: Also supports practitioners to embed community practice locally.

Collaboration: Also engages with the design, development and review of community practice. Consults and collaborates widely, potentially taking a community coordination and leadership role. Driving maintenance and updates of existing practice and identifying new areas for the development of policy and implementation standards. Actively communicating and promoting existing and emerging approaches to the immediately impacted communities and the wider data infrastructure landscape

[26] https://cmmiinstitute.com/learning/appraisals/levels

Appendix 4: Diagram 5: CoreTrustSeal Requirements to FAIR Principles Alignment (large)