FAIRSFAIR
Fostering Fair Data Practices in Europe

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

# M4.1 Evaluation of current CoreTrustSeal guidelines and extended guidance to consider their implications for maturity modeling

| | |
|---|---|
| Work Package | WP4 FAIR-Certification |
| Lead Author (Org) | Hervé L'Hours (UKDA) |
| Contributing Author(s) (Org) | Ilona von Stein, Jerry de Vries, Frans Huigen, Mustapha Mokrane (DANS), Anusuriya Devaraju (UniHB), Joy Davidson, Patricia Herterich (DCC) |
| Due Date | 31.03.2020 |
| Date | 31.03.2020 |
| Version | 1.0 |

| | |
|---|---|
| X | PU: Public |
| | PP: Restricted to other programme participants (including the Commission) |
| | RE: Restricted to a group specified by the consortium (including the Commission) |
| | CO: Confidential, only for members of the consortium (including the Commission) |

# Introduction & Overview

This paper is milestone 4.1 of the FAIRsFAIR task 4.1 (Capability Maturity models towards FAIR Certification) within the FAIR Certification work package (WP4).

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

This document presents the first iterative step in aligning the characteristics of FAIR digital objects with the repositories that 'enable' FAIRness, through the CoreTrustSeal Trustworthy Data Repository Requirements[1] and the application of a capability/maturity evaluation approach.

The CoreTrustSeal Requirements, assessment process, and governance are a community-driven effort to identify best practices, support improvement, and deliver better repository service outcomes to data users. Certification offers recognition and demonstrates trustworthiness to data depositors, users and funders. But it is through the process of self-assessment and peer review that practices are shared and data infrastructures are improved. This FAIRsFAIR process follows that spirit of open inclusivity. The goal is to share and improve rather than exclude repositories or digital objects. Gaps in trustworthy repository practice or FAIR objects' status are opportunities for discussion and targeted improvement.

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

The selection and application of capability and maturity levels to processes, activity areas, and organisations is the starting point in the design of standard requirements and assessment processes for CoreTrustSeal and FAIR. The challenge is to develop an approach which has operational value and is sustainable.

This is an initial evaluation of the current CoreTrustSeal Requirements extended guidance to consider their implications for maturity modelling in the context of FAIR. We refer to this as CoreTrustSeal+FAIR. It will be iteratively updated to support the evaluation of Trustworthy Digital Repositories (TDR), including their ability to offer an environment which enables FAIR data and metadata for the long term.

[1] https://zenodo.org/record/3632533

The context surrounding the work package, project, FAIR data and trustworthy digital repositories is briefly described. The methodology of the approach is explained and the design principles of the proposed approach are outlined, including some issues and dependencies.

The conclusion and next steps explain how the proposed approach will be opened to initial feedback and testing before a round of iterative updates.

# Scope & Context

The primary focus of this work is to align CoreTrustSeal Requirements with FAIR to identify how repositories can enable FAIR data. Provision of a capability maturity approach is central to this work, but the application of capability and maturity levels will not be prescriptive at this stage. These will be developed iteratively through interaction with ten supported repositories and through wider engagement, including the emerging European Network of Trustworthy Digital Repositories enabling FAIR data.

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

**The FAIR Data Principles: Baseline**
The detailed clarification of each principle and its application is beyond the direct scope here, though highly relevant to any final recommendations.

All current FAIR work can be traced back to the original 2014 Force 11 Principles and the subsequent Nature paper[2] which we use as our reference point. The numerous ongoing efforts around FAIR often question the meaning and intention of the original principles at different points in their work. We need to address these issues of FAIR interpretation without allowing them to delay our progress. We have annotated the original principles to develop a 'baseline' of potential issues[3] that would impact defining and evaluating object FAIRness or the ability of repositories to enable FAIRness.

These baseline issues are used as a reference point in each stage of developing FAIR-related work. Each iteration should either address the baseline issues, or acknowledge that they have not been addressed.

[2] https://www.nature.com/articles/sdata201618
[3] https://zenodo.org/record/3728131 FAIR Principles: Baseline Comments

**Repository Interoperability**
As components of the EOSC the interoperability of the repositories themselves is important. This particularly applies to technical standards for repository interoperability. Full details of the FAIRsFAIR work in this area are presented in *D2.3 Set of FAIR data repositories features*[4]. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

**Object Assessment**
Among the many moving targets in FAIR and EOSC is the agreement of indicators and tests for objects' compliance with the FAIR principles. Full details of the FAIRsFAIR work in this area, including interactions with the RDA FAIR Data Maturity Working Group are available in the deliverable *4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories*[5]. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

**Service Assessment**
Repositories are part of a wider data service infrastructructure. Full details of the FAIRsFAIR work in this area are available in *Assessment Report on FAIRness of Services*'[6]. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

**Wider EOSC Components**
Repository interactions with and dependencies upon the wider components of a distributed research data ecosystem such as the EOSC will be critical to the final recommendations from this work. FAIR Ecosystem Components: Vision[7] has been shared and will be iterated in response to external feedback and internal finding. We will engage with this work and outcomes will be integrated into future iterations of CoreTrustSeal+FAIR.

**Assessment & Evaluation Modelling**
The outcome of an assessment and evaluation process is a statement of the status of an object or entity (e.g. Trustworthy, FAIR, Open). With an extensive range of disparate evaluation approaches in operation or in development it's helpful to develop a structured typology of concepts and how they interact. This lets us design and evaluate evaluation standards and processes and compare them. We have developed a generic assessment and evaluation reference model[8]. Future iterations of the CoreTrustSeal+FAIR outcomes will be benchmarked against this model.

[4] https://zenodo.org/record/3631528
[5] https://zenodo.org/record/3678716
[6] https://zenodo.org/record/3688762
[7] https://zenodo.org/record/3734273
[8] https://zenodo.org/record/3733280

# FAIR Objects & FAIR Enabling Environments

Different repositories work with different assumptions about what is a 'digital object' and how the content of an object is divided into 'data' and 'metadata'. In T*urning FAIR Data into Reality*[9], the following overview object model is presented.

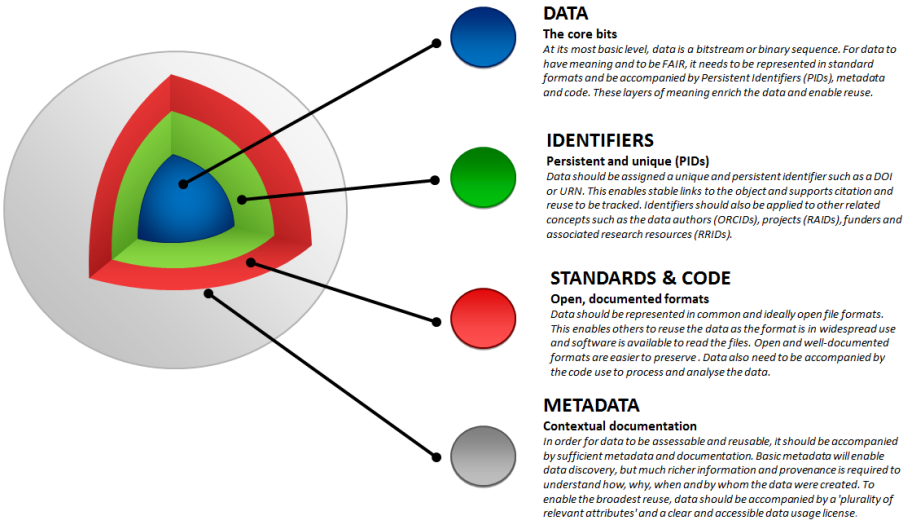


Diagram: **Rec. 3: A model for FAIR Data Objects**

But this division between the data (as the original target for collection/creation) and its supporting metadata is not always as clear and consistent in reality. For example under a standard like DDI[10] data and associated metadata may be contained within a single file. Repositories also create their own organisational (meta) data while administering the digital objects which we'll refer to as 'business information' to differentiate it from the digital objects' data. This business information may include policies, procedures and workflows, and may have it's own 'metadata' (ranging from 'policy review/approval' to 'format migration quality check result). Some of this repository 'process' metadata (e.g. 'validation of a checksum' or 'format risk assessment outcome') might be stored and managed with the object metadata. All of these (meta) data types are important as they are either our target for ensuring FAIRness or they provide supporting evidence for enabling FAIRness.

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

[9] https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[10] https://ddialliance.org/Specification/

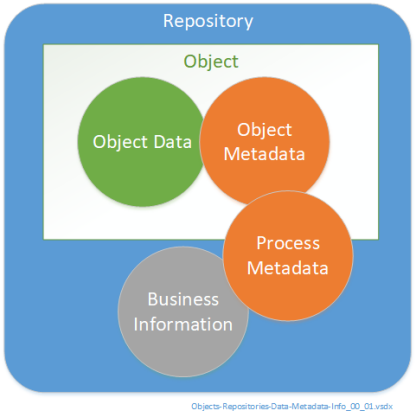


Diagram: **Repository & Object Metadata**

In the development and implementation of CoreTrustSeal+FAIR we must take account of repositories' and their collections of heterogeneous digital objects but make the practical decisions needed for a broadly applicable standard approach.

The diagram below demonstrates a mapping from objects to the FAIR principles which takes account of the (repository) context and some wider dependencies.

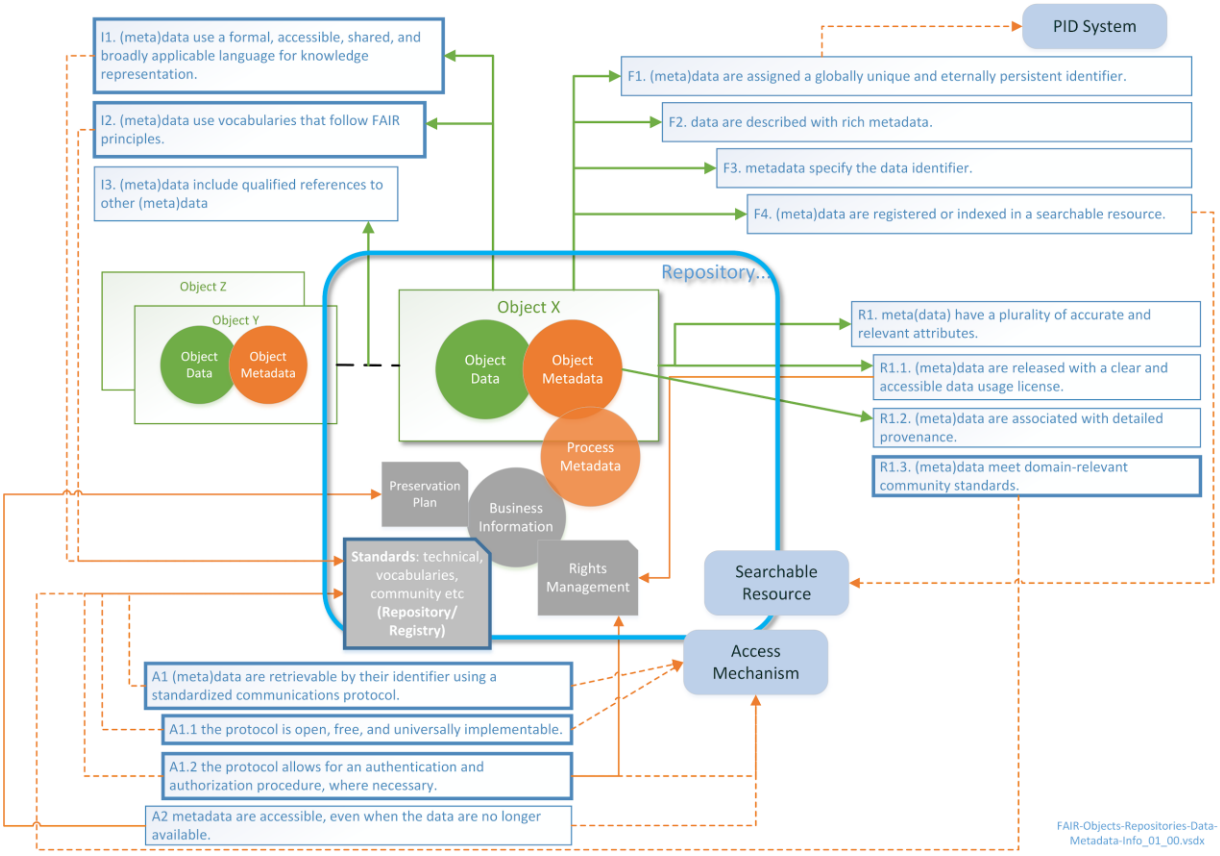


Diagram: **FAIR Objects Repositories, Dependencies** (FAIR Principles abbreviated)

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

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

Identifying these potential dependencies is important to defining the alignment between objects and their repository environment and more broadly to identifying other actors which may provide supporting evidence for CoreTrustSeal$^{+}$FAIR status.

## CoreTrustSeal Requirements in Brief

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

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

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

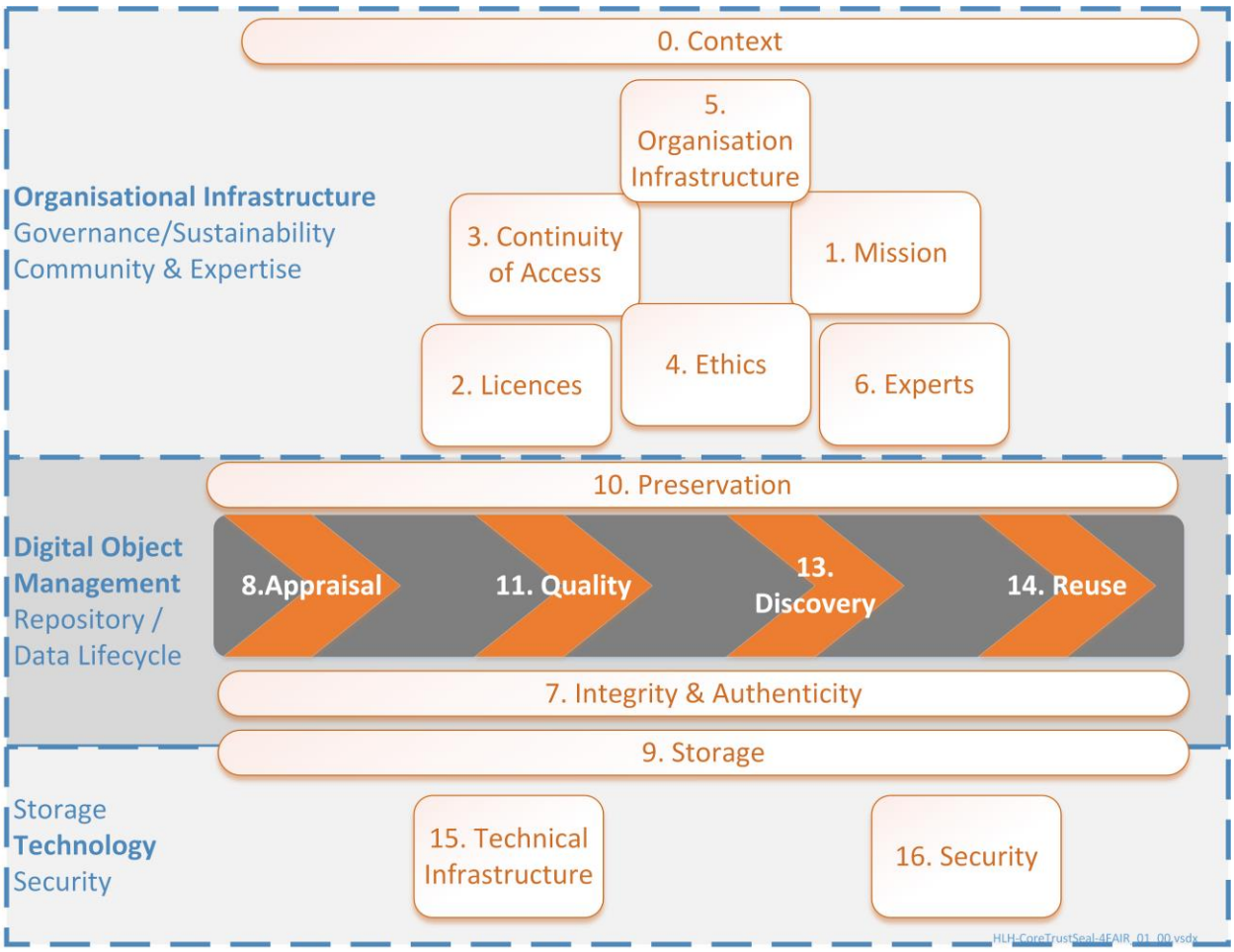


Diagram: **CoreTrustSeal Requirements in brief**

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

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

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

# CoreTrustSeal+FAIR

The first iteration of the CoreTrustSeal to FAIR alignment mapping is currently being reviewed and responded to by the ten  FAIRsFAIR Repositories supported within this FAIRsFAIR work package. The *CoreTrustSeal+FAIR Overview*[11] presents a number of high level FAIR-related questions, asks for additional repository context, and maps the FAIR principles and the indicators being evolved by the FAIR Data Maturity Working Group[12] to the CoreTrustSeal Requirements.  We have a number of areas where the requirements can be aligned directly with repository capability, but a challenge that there are multiple areas of repository activity where FAIR might be assured (e.g. Appraisal, Quality and Reuse).

The outcome of this feedback and review process will be a FAIR mapping integrated into a template of the CoreTrustSeal Extended Guidance.

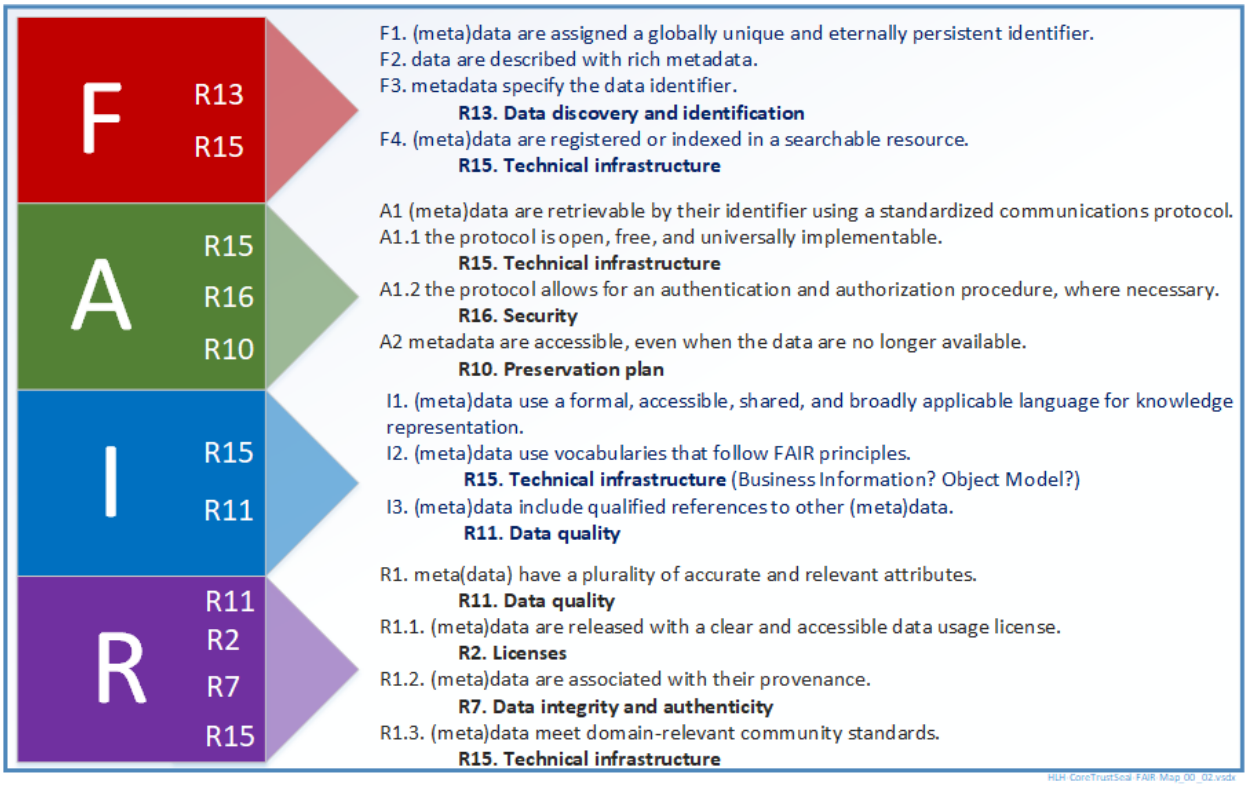


Diagram: **FAIR to CoreTrustSeal**

[11] https://zenodo.org/record/3734897
[12] https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

# CoreTrustSeal+FAIR: Draft Elaboration Model

In setting up an approach for FAIR enabled repositories we need to consider where we can elaborate on the existing CoreTrustSeal requirements and whether some additional features are required.

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

The design methodology is to use the CoreTrustSeal Requirements as a baseline and to elaborate them in ways which demonstrate that a repository enables FAIRness. We will consider the implications of the fact that neither the FAIR principles, nor the CoreTrustSeal criteria were developed with a focus on 'capability' or 'process evaluation' for maturity assessment.

The repositories selected by FAIRsFAIR are the primary audience for development, implementation and iteration, but much wider feedback is sought as we iterate and test the approach. There are a number of logical mappings from FAIR into various parts of the Requirements. But we need to select the most intuitive and practical mapping so repositories have clear locations to provide evidence statements and associated evidence for FAIR enabling.

The direct mapping of FAIR and CoreTrustSeal and the application of capability and maturity assessments has a number of challenges. The FAIR acronym is expanded into 15 principles, each of which is under review to develop relevant indicators, metrics, and tests. The RDA FAIR Data Maturity Working Group are also classifying each indicator as one of: essential, important or useful. At this stage the working group does not have agreed indicators, metrics and tests to assess compliance with every principle.

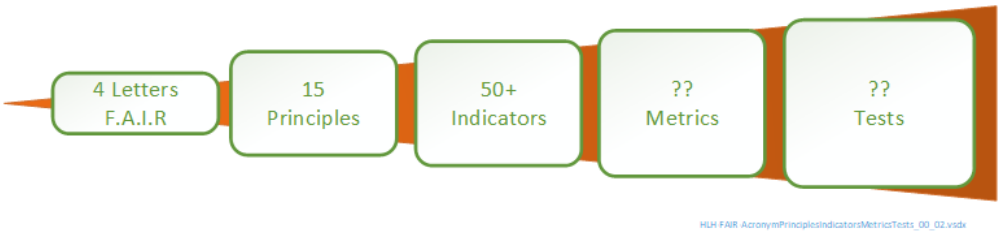


Diagram: **FAIR acronym, principles, indicators, metrics & tests.**

For the initial integration of CoreTrustSeal assessment with the FAIR principles we have a number of open issues. These include the need for feedback from repositories about their perception of FAIR enabling and a wider set of contextual questions than those currently requested by CoreTrustSeal. The implications of applying capability and maturity levels are discussed below. Some CoreTrustSeal requirements may need to be divided into more granular

capability assessments. There are also some FAIR concepts, including the use of standards and the provision of access functionality, which are implied by several CoreTrustSeal Requirements rather than being explicitly defined.

## CoreTrustSeal, Compliance, Capability & Maturity

Those self-assessing against the CoreTrustSeal Requirements are provided with five tiers of compliance for their responses:

**CoreTrustSeal 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 compliance levels have some alignment to maturity thinking. Though level 0 (not applicable) is arguably not on the same scale as 1-4. The supporting guidance for the compliance levels states:

> "Compliance levels are an indicator of the applicant's self-assessed progress, but reviewers judge compliance against response statements and supporting evidence. If an applicant believes a Requirement is not applicable (0), then this must be justified in detail. 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)."

In theory a capability/maturity measure may be applied to any defined context (e.g. a repository) for any defined set of processes and outcomes. Capability/maturity levels are one of many evaluation scales, and even the range of similar maturity scales in use can be challenging to manage and implement.

The initial FAIRsFAIR work on CoreTrustSeal maturity mapping works through the current standard Capability Maturity Model Integration CMMI[13] (levels of capability and performance[14]).

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

But these aren't fully aligned with the prior CMM Software[15] work

[13] https://cmmiinstitute.com/
[14] https://cmmiinstitute.com/learning/appraisals/levels
[15] https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=11955

| 1: Initial | 2: Repeatable | 3: Defined | 4: Managed (Capable) | 5: Optimizing (Efficient) |
|---|---|---|---|---|

Though complex mapping and evaluation exercises are necessary to deliver a CoreTrustSeal/FAIR maturity alignment, the resultant standard and process must be practical, implementable and as simple as possible. In contrast to the granular maturity assessment of a particular process as "5: Optimizing" (CMMI), the CoreTrustSeal makes a broader assumption about repository progress over time.

Supported repositories will evaluate themselves against the CoreTrustSeal compliance levels and support the iterative development of appropriate capability/maturity expectations. Trustworthy digital repository standards, including CoreTrustSeal, were not developed with maturity in mind. Though maturity assessments could be applied across the requirements we must consider practicality for implementation. CoreTrustSeal is not only 'core' because it seeks to cover all the basic TDR requirements. It is 'core' because it tries to retain a level of structural simplicity and usability:
`Requirement > Guidance > Evidence Statement > Evidence Links.`

Examples:

Example: Mission (R1) may be mapped relatively easily to the maturity levels of the management and approvals process of a "mission statement" document. But repositories exist in a range of organisational structures (standalone organisations, departments in universities, partnerships) mean that applicants do not always have control over the full organisational mission. In this case the human evaluation of evidence and evidence statements can provide a more nuanced assessment than a strict assignment of maturity levels.

Example: Data integrity and authenticity (R7) are separate concepts but they are codependent indicators of trustworthy practice in that unintentional change must be avoided and intentional change documented. The CoreTrustSeal process can provide an overall assessment, but a maturity approach might require separate assessments for integrity and authenticity.

The CMMI levels represent a starting point for designing a tiered CoreTrustSeal+FAIR capability/maturity approach. Levels 0 to 3 align well with standards like FitSM[16] and we will examine how '4. Quantitatively Managed' aligns with the goal of improving automation and machine-actionability of processes and objects. Tiers 4 and '5. Optimizing' are a high bar. Tiers should be seen as an initial basis for discussion, there is no pass/fail outcome implied.

Capability/Maturity tiers may be usefully applied to both evidence management (including artefacts like mission, preservation plan, technical infrastructure etc.) and workflows (including for appraisal, quality assurance, re-use).

[16] https://www.fitsm.eu/downloads/#toggle-id-7

Later outcomes might include identifying ‘minimal’ or ‘ideal’ levels for Trust or FAIR criteria. Based on testing the tiers may be simplified and recommendations made to CoreTrustSeal for capability/maturity adoption in their next revision process.

## Capability & Maturity Levels: Overview & Implications

The Italicised text below is taken from: https://cmmiinstitute.com/learning/appraisals/levels. Notes below each section are intended to support discussion about how this draft elaboration model should apply the capability and maturity tiers.

The goal is to design an approach that enables the FAIRness of digital objects while retaining the CoreTrustSeal low-barrier-to-entry, ‘core’ approach. The CMMI evaluation approach is relatively complex and detailed and the overall standard is transitioning from version 1.3 to version 2.0 (September 2020). For at least the initial FAIRsFAIR iteration we will concentrate on capability/maturity assessment without integrating the CMMI capability areas and practice areas.

> ***Capability Levels***
> *Capability levels apply to an organization’s performance and process improvement achievements in individual practice areas. Within practice areas, the practices are organized into practice groups labeled Level 0 to Level 5 which provide an evolutionary path to performance improvement. Each level builds on the previous levels by adding new functionality or rigor resulting in increased capability.*

The initial mapping of CoreTrustSeal to FAIR will either treat Requirements as one or more practice areas, or will integrate additional areas of practice into the R0: Context section of CoreTrustSeal. For each area we will seek a self-assessment against levels 1-3 and ask self-assessors to consider what might be required at the practice level to support level 4 or 5 maturity at the organisation level.

> ***Maturity Levels***
> *Maturity levels represent a staged path for an organization’s performance and process improvement efforts based on predefined sets of practice areas. Within each maturity level, the predefined set of PA’s also provide a path to performance improvement. Each maturity level builds on the previous maturity levels by adding new functionality or rigor.*

Maturity levels will be generated based on a calculation of capability levels outcomes and their integration repository-wide practice. Capability and maturity levels 0 to 5 are presented together below for comparison.

> ***Capability Level 0: Incomplete***
> *Incomplete approach to meeting the intent of the Practice Area.*
> *May or may not be meeting the intent of any practice.*
> *Inconsistent performance.*

***Maturity Level 0: Incomplete***
*Ad hoc and unknown. Work may or may not get completed.*

Any self-assessments at level 0 will be reviewed and prioritised.

***Capability Level 1: Initial***
*Initial approach to meeting the intent of the Practice Area.*
*Not a complete set of practices to meeting the full intent of the Practice Area.*
*Addresses performance issues.*

***Maturity Level 1: Initial***
*Unpredictable and reactive. Work gets completed but is often delayed and over budget.*

Any self-assessments at level 1 will be reviewed and specific guidance developed.

***Capability Level 2: Managed***
*Subsumes level 1 practices.*
*Simple, but complete set of practices that address the full intent of the Practice Area.*
*Does not require the use of the organizational assets.*
*Identifies and monitors progress towards project performance objectives.*

***Maturity Level 2: Managed***
*Managed on the project level. Projects are planned, performed, measured, and controlled.*

Our initial assumption is that level of 2: Managed should be the targeted minimum across the self-assessments at the end of the support process.

***Capability Level 3: Defined***
*Builds on level 2 practices.*
*Uses organizational standards and tailoring to address project and work characteristics.*
*Projects use and contribute to organization assets.*
*Focuses on achieving both project and organizational performance objectives.*

***Maturity Level 3: Defined***
*Proactive, rather than reactive. Organization-wide standards provide guidance across projects, programs, and portfolios.*

At level three and above the practice areas are managed as part of repository-wide practice. We will evaluate whether ‘defined’ should be a minimal level for any capability area. Beyond the CoreTrustSeal+FAIR work a ‘defined’ level of practice may be necessary to support aspects of organisational interoperability, e.g. to contribute to the European Open Science Cloud (EOSC).

***Maturity 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.*

Under CMMI 'Quantitatively managed' is a necessary precursor to 'optimizing'. For each Requirement we will ask for participant feedback on what they would see as necessary to become quantitatively managed. This will be used to develop criteria for level 4 maturity at the organisational level. Levels 4 and 5 below are both acknowledged as a high bar for some organisations and one which might depend on local circumstances (e.g. relationship with a host organisation). FAIRsFAIR will consider whether level 4 is a necessary precursor to the machine-actionability and automation envisaged by the creators of the FAIR principles.

***Maturity 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.*

While 'optimizing' is a desirable goal, it is not without significant resource implications at the organisational level. There is not yet a clear case where demonstrating an 'optimizing' level of maturity is required for CoreTrustSeal, FAIR, or EOSC, though the latter may be addressed by efforts to assess FAIR services. For each Requirement we will ask for participant feedback on what they would see as necessary to reach this level, this will be used to develop criteria for level 5 maturity at the organisational level.

## Assessment Methods & Outcomes

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

The iterative self-assessment process supports defining a final assessment method which will result in agreed outcomes including the defined 'status' of a repository e.g. as CoreTrustSeal$^{+}$FAIR enabling. This work takes place in parallel to efforts to test and 'badge' individual digital objects as 'FAIR'.

Repository support in FAIRsFAIR will enable applications for CoreTrustSeal which integrate evidence for FAIR enabling, but during the course of this work there is no 'pass/fail' outcome within FAIRsFAIR or formal process of FAIR enabled certification through CoreTrustSeal. Recommendations for integration will be shared and discussed with the CoreTrustSeal Board.

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

## Certification and Badging

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

## Change, Periodicity & Validity Terms

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

# Open Issues for Integration

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

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

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

**Best, Minimal and Ideal Practices**
The existence of standards like CoreTrustSeal, OAIS, ISO16363, ISO27001 and others does not mean there is always a community consensus on minimal levels of service and necessary supporting evidence. The CoreTrustSeal is the only current effort generating a publicly available body of work which could be used to support discussion on the often used phrase 'best practices'. For formal assessment of object or repository characteristics it's necessary to move from general assumptions of what 'best practice' means to SMART (specific,

measurable, achievable, realistic, time bound) objectives. We might also usefully differentiate between ‘minimal practice’ and ‘ideal practice’. Some levels of practice might be defined purely from a “technical perspective” e.g. a minimal number of data copies, while others will be dependent on local context including the needs of the data users.

**Designated Community & other Users**
We will seek clearer approaches to defining designated communities, and agreement on expectations of how a repository should interact and respond to their needs.

“Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time”[17].

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

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

**The Full (FAIR) Data Lifecycle & Ecosystem**
In line with the wider vision for FAIR the FAIR-enabling repository work must integrate and align with a vision of the full FAIR data ecosystem and data lifecycle. This includes identifying how to align with work on research data management plans.

**Non-(Meta) Data Artefacts as Evidence**
For any evaluation of FAIRness other than direct inspection of an entity (individual review of an object, site visit to a repository etc.) there is some dependence on the provision of evidence. Evidence could range from mission statements, policies, procedures and workflows, to granular outcomes of fixity checks. This evidence is another type of ‘digital object’ generated as a result of running any infrastructure (people, processes, technology) which curates digital objects.

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

[17] https://zenodo.org/record/3632563

# Conclusions and Next Steps

At this stage of the iterative process we have a draft alignment between the FAIR Principles and the CoreTrustSeal Requirements. We have outlined the capability and maturity approach which will be applied to the CoreTrustSeal+FAIR alignment. The responses to our high level FAIR questions to repositories will help validate and improve the alignment. Our upcoming deliverable will take the first steps to guide repositories in self-assigning capability levels. These self-assigned levels will be used to develop further recommendations on tiered capability definitions for CoreTrustSeal+FAIR. As agreed capability measures emerge the calculation of overall repository maturity will be addressed.

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

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

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

# Appendix: FAIR Objects, Repositories, Dependencies