



COAR Framework: CoreTrustSeal Board Feedback

The CoreTrustSeal Board¹ thanks the COAR for this opportunity to comment during the review of the 45 Criteria of the v1.0 COAR Community Framework for Good Practices in Repositories (Public version 1 – October 8, 2020²). The general and detailed specific feedback below has also been added to the COAR online feedback form.

We are glad to see ongoing work in this area, and have cooperated successfully with COAR in the past. The CoreTrustSeal provides requirements and a supporting process to deliver the need, as defined by the Research Data Alliance (RDA)³, of providing a single reference point for 'Core' Trustworthy Digital Repository (TDR) Requirements. Our general feedback primarily relates to concerns if the community is faced with guidance that lacks clarity, is repetitive of existing material, or where there is a risk of divergence over time.

You mention *"a number of existing frameworks and evaluation criteria that were developed to assist repositories in assessing certain facets of their operations (such as discovery, access, reuse, integrity, quality assurance, preservation, privacy, and sustainability),"* but note that *"these criteria are spread across different organizations and are often relevant for only one region or one type of repository"*⁴ and then state that *"The following frameworks [including CoreTrustSeal] were taken into consideration for this work"*. The relationship between the COAR criteria and these sources is not made explicit, which risks the COAR approach becoming another of *"a number of existing frameworks and evaluation criteria"* and one which may diverge from these sources as they grow and change over time. One example might be the FAIR Data Principles, where the associated RDA FAIR Data Maturity Indicators have provided greater specificity since the COAR criteria were released.

The CoreTrustSeal has previously discussed providing guidance on how more specific (technical, disciplinary or otherwise more rigorous) expectations could be built around the 'Core' of CoreTrustSeal. We will seek to progress this work and to cooperate with COAR in that process.

One of the types of input requested is *"Links to examples, guidance, and best practices for implementation of a given criteria in repositories"*, we suggest that these should then be used to iteratively review the structure and content of the criteria. The questions relating to *"Ease of Adoption – Is this characteristic too difficult to implement"* and *"Relevance – Is criteria appropriate for all repositories?"* raise the question of who the COAR is seeking to include and exclude from this framework. The open consultation refers to the 'COAR Repository Assessment Working Group'

¹ <https://www.coretrustseal.org/about/standards-and-certification-board/>

² <https://doi.org/10.5281/zenodo.4110829>

³ <https://www.rd-alliance.org/>

⁴ <https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/>

that produced the version 1.0, but does not provide further details, or define how the mission and scope of COAR itself influences their perspective on these requirements.

It might also be helpful to explain the decisions taken about 'essential' and 'desirable' including the risks and outcomes if these are not met. We hope that representatives of COAR will in turn find time to provide feedback revision on the CoreTrustSeal Requirements:. We'd be particularly interested to understand if, and why, there are any CoreTrustSeal Requirements that are not necessary for the COAR criteria.

Feedback Detail

OAR Community Framework for Good Practices in Repositories. Reference: <https://doi.org/10.5281/zenodo.4110829>

Objective	Essential / Desired	Number	Description	CoreTrustSeal Board Feedback
1. Discoverability	Essential	1.1	The repository supports quality metadata and controlled vocabularies (discipline-based, regional or general metadata schema such as Dublin Core)	CoreTrustSeal has to date avoided a separate 'Metadata' section because different types of metadata are implied across repository activities. This criterion is very specific to 'Discoverability' but the need for "discipline-based, regional or general metadata" is much more broad.
	Essential	1.2	The repository supports harvesting of metadata using OAI-PMH	A common approach, but a very specific technical requirement and potentially exclusionary for some, depending on the scope of the criteria
	Essential	1.3	The metadata in the repository are available, even in cases when the resource is no longer available	Good practice, though this could directly reference the relevant FAIR Data Principle
	Essential	1.4	The repository assigns a persistent identifier (PID) that points to the landing page of the resource, even in cases where the resource is not available	This is a very specific requirement that a 'landing page' (undefined) be used. Landing pages vary in structure and content and can present a barrier to machine-accessible metadata and data and this approach excludes direct resolution to a 'digital object'. The continued resolution to metadata after a resource ceases to exist is already covered in 1.3.
	Essential	1.5	The repository offers a search facility and/or the metadata is indexed by external discovery services and/or aggregators	Good practice, though this could directly reference the relevant FAIR Data Principle

	Essential	1.6	The repository is included in one or more disciplinary or general registries of resources (e.g. Re3data, OpenDOAR or other national, regional or domain registries)	This relates to the discoverability of the repository rather than the digital objects it contains.
	Desired	1.7	The repository supports linking between related resources such as articles, data and software (e.g. including PIDs for related resources held elsewhere)	Good practice, though this could directly reference the relevant FAIR Data Principle. The repository supporting this is different from the repository doing this or ensuring it is done.
	Desired	1.8	The repository supports HTTP link headers to provide automated discovery of metadata records and content resources associated with repository items. We recommend Signposting typed links to support this.	A very specific and potentially exclusionary technical requirement, depending on the scope of the criteria. Should probably provide a link to https://signposting.org/
	Desired	1.9	The repository supports PIDs for authors, funders, funding programmes and grants, institutions, and other relevant entities	PIDs for as wide a range of entities as possible are certainly desirable.
	Desired	1.10	The metadata in the repository are available under a Creative Commons Public Domain License and are available for download in a standard bibliographic format	This is ideal. but is a very specific and potentially exclusionary requirement, depending on the scope of the criteria. Not all metadata curators have this license option due to complex IP in the sourcing of metadata
	Desired	1.11	The metadata in the repositories are available in human-readable and machine-readable formats	Obviously desirable, though would benefit from further definition of machine readable
	Desired	1.12	In the case of data, the repository supports PIDs for data at multiple levels of granularity, where appropriate (for example, if there there is research using a subset of the full dataset, a citation of the data subset will be needed)	Does this require the adoption of an (unspecified) formal PID system. Or would locally unique and persistent identifiers within an object with a PID be acceptable?
2. Access	Essential	2.1	There is no cost to the user for accessing the resources in the repository	A very specific and potentially exclusionary requirement, depending on the scope of the criteria. Is this a requirement for COAR membership? The criteria do not explicitly mention open access either do they?

	Essential	2.2	The repository ensures ongoing access to resources for a publicly stated time frame	Clear minimum retention periods are valuable. Should the outcomes of reappraisal (e.g. a change to retention or to curation/preservation level) after that period also be transparent?
	Essential	2.3	The repository supports access to its documentation and metadata for persons with disabilities	Improved accessibility for all is something we must support and deliver.
	Essential	2.4	Device neutrality – no specific device needed for users to access the repository	This probably needs expanding. user-owned device? Or would this cover systems that provide data visualization or secure remote access?
	Desired	2.5	The repository provides a mechanism to make very large files available to users outside of the normal user-interface (in cases where the size of the file becomes unwieldy for the user)	This is coming from a very 'data file' and 'download' driven perspective.
	Desired	2.6	In cases where there is restricted access to a resource, the repository facilitates an indirect way to access this resource (e.g. by contacting the author)	If a repository does not provide a method for accessing a resource then it is not accessible
				2.6 implies the need for access rights too. And a repository must have sufficient rights to store, curate and preserve too.
3. Reuse	Essential	3.1	The repository includes licensing information in the metadata record which stipulates reuse conditions	
	Essential	3.2	The repository provides citable PIDs (see 1.4)	
	Essential	3.3	The repository provides a list of preferred, non-proprietary formats	Non-proprietary may be desirable, but there may not be options and proprietary may also be open. Preferred for what, immediate reuse? There is no equivalent criterion under 'Preservation' below.
	Essential	3.4	The landing pages include the metadata about the item including information required for citation in machine and human readable format	See comments on landing pages. Requiring that there is a link from the digital object to this information would be more generally applicable
	Desired	3.5	The repository has open APIs to support full text harvesting and/or text and data mining	A common approach, but very specific and potentially exclusionary, depending on the

				scope of the criteria
	Desired	3.6	The resources are stored in machine-readable, community standard formats	Examples of non-machine readable formats? Source of defined community standards? Overlap with 3.3
4. Integrity and authenticity	Essential	4.1	The repository provides documentation or has a policy outlining its practices that prevent unauthorized access/manipulation of resources	
	Essential	4.2	The repository keeps a record of all changes to metadata and resources in the repository	
	Essential	4.3	The repository supports versioning of metadata and resources after deposit	
	Desired	4.4	The repository provides information about the content provider(s) in the metadata including the name of the person(s) and/or institution(s) responsible for the resource	Responsible?
5. Quality Assurance	Essential	5.1	The repository undertakes basic curation of metadata (and resources, if applicable) i.e. brief checking and addition of basic metadata or documentation where needed.	"brief checking and addition of basic metadata or documentation where needed" is taken from the CoreTrustSeal levels of curation. But what is the COAR approach to unattended self-deposit systems?
	Essential	5.2	The repository provides documentation or has a policy outlining what curation processes are applied to the resources and metadata	
	Desired	5.3	The repository supports external annotation, commenting or reviewing of resources and metadata	
6. Privacy of sensitive data	Essential	6.1	In cases where the repository is collecting sensitive research data, there are mechanisms that allow data owners to limit access to authorized users only	Overlap with 4.2
	Desired	6.2	In cases where the repository is collecting sensitive research data, the repository provides tiered access based on the different levels of security requirements of data	This seems to say "if there are different access restrictions you should apply them?" overlap with 4.2 and 6.1

7. Preservation	Essential	7.1	The repository (or organization that manages repository) has a long term plan for managing and funding the repository	
	Essential	7.2	The repository provides documentation or has a policy that defines the duration of time the resources will be managed over the long term and documentation about preservation practices	This conflates retention periods with minimum guaranteed periods where active preservation will be undertaken
	Desired	7.3	Repository has a documented approach to preservation, that adopts widely accepted preservation practice	"widely accepted" may need some supporting references and additional detail.
	Desired	7.4	The agreement between depositor and repository provides for all actions necessary to meet preservation responsibilities - e.g. rights to copy, transform, and store the items	Cf ReUse rights separately under 4.2
8. Sustainability and governance	Essential	8.1	The repository clearly indicates what organization is responsible for managing the repository	
	Essential	8.2	The repository clearly indicates the nature of the governance of the services (or the organization that manages the repository)	
9. Other	Essential	9.1	The repository has a contact point or helpdesk to assist depositors and users	
	Essential	9.2	The repository has a public notice stating it will respond to queries within a certain time frame (which is no longer than 14 days).	This is oddly specific compared to the other criteria
	Essential	9.3	The repository provides documentation or has a policy that outlines the scope of content accepted into the repository	
	Essential	9.4	The repository collects and shares usage information using a standard methodology (e.g. number of views, downloads)	This is logical and desirable, but what is the reason for it being essential when, for example. 9.5 is Desired.
	Desired	9.5	The repository functions on well-supported operating systems and other core infrastructural software	
	Desired	9.6	The submission / deposit system supports both individual creator uploads and bulk uploads of records and resources.	Doesn't this depend on the type and scale of the repository and digital objects?