Checklist for Evaluation of Dataset Fitness for Use

For use by repository managers or a similar external entity to evaluate the data holdings of repository for fitness for use. Answer these questions for a sample (~6 to ~12) of datasets in that repository.

This checklist is meant to supplement the CoreTrustSeal Repository Certification process. They are based on the criteria stated in Fitness for use: compilation of criteria under “CoreTrustSeal extension: evaluation of a sample (6-12) of individual data sets” (Column E). See "https://www.rd-alliance.org/system/files/Fitness%20for%20use_%20compilation%20of%20criteria_0.xlsx" (You'll need to cut and paste this link into a web browser)

If the dataset to be evaluated is not stewarded within a CoreTrustSeal-certified repository, this checklist will not be useful in evaluating dataset fitness for use.

There are three categories for fitness:
“does not meet this criteria for fitness”, or “No/N”
“somewhat meets this criteria for fitness”, or “Somewhat/S”
“meets this criteria for fitness” or “Yes/Y”

A set of asterisks (***) denotes that assessing this criteria will require domain- or discipline-specific knowledge.

Caveats:

This checklist is for assessment of a dataset’s fitness for use prior to accumulation of usage metrics. Usage metrics are useful in evaluating a dataset's fitness for use for repositories of sufficient age.
Use of this checklist for assessment of a dataset’s fitness for use does not necessarily provide for an assessment of the fitness for use for all files in the dataset. The fitness of use for any file within a dataset may vary from the fitness of the dataset.
Currently this checklist is not geared towards all forms of scholarship and would need further generalizing to a wider audience.

For a copy of the full checklist prior to Google form usage go to "https://www.rd-alliance.org/system/files/ChecklistForEvaluationOfDatasetFitnessforUse_v2_20181218_0.pdf" for a PDF version of the questions. (You'll need to cut and paste this link into a web browser)

Additions to this checklist should include:

A rating system for each dimension (e.g. metadata completeness, accessibility) that aligns with the CoreTrustSeal certification process
Additional guidance on criteria in an appendix, and possibly visible through mouseovers.

* Required

Dataset and Assessor Identification

1. **Repository**

2. **Location of Dataset (PID or URI preferred)** *

3. **Assessor's Name and e-mail address**

4. **Originating Domain of Dataset (include subdomain if app**

Repository Certification

5. **Dataset is located within a CoreTrustSeal-certified repository** *
Mark only one oval.

- Yes
- No

6. **Dataset is located within a World Data System or Data Seal of Approval certified repository (see https://www.coretrustseal.org/why-certification/certified-repositories/ for a list)** *
Mark only one oval.

- Yes
- No

Curation

7. **Repository representative stipulates that structure, harmonization, completeness, and correctness of the dataset comports with typical data curation activities conducted by this (CoreTrustSeal certified) repository**
Mark only one oval.

- Yes
- Somewhat
- No

8. *Mark only one oval.*

- Option 1

Metadata Completeness

9. **Citation exists, including authorship, year, comprehensive title, persistent identifier (e.g. DOI)**
Mark only one oval.

- Yes
- Somewhat
- No

10. ***Description of the dataset content adequate; this includes the following:**
Check all that apply.

	Yes	Somewhat	No
An abstract	☐	☐	☐
A listing of measurement & observation types, or parameterizations & simulation types including methods used	☐	☐	☐
a description of size, structure, and data format/MIME type	☐	☐	☐

11. **Coverage (spatial, temporal, or other dimensions) adequate**
Mark only one oval.

- Yes
- Somewhat
- No

12. **Description of provenance, this includes:**
Check all that apply.

	Yes	Somewhat	No
authorship/contributorship (data creators and sources, institutions and laboratories involved in data generation)			
a description of data collection or generation (sampling events, used devices, processing steps, QA/QC, experimental or model setup including parameterization)			
References to related work (scholarly work using the data, method papers, references to preceding or related data products)			

13. ***Additional metadata adequate to respective research domain (if applicable)**
Mark only one oval.

- Yes
- Somewhat
- No

14. **Terms of usage (licenses, other conditions of reuse, data protection, ethical issues)**
Mark only one oval.

- Yes
- Somewhat
- No

Accessibility

15. **Persistent identification of the dataset and related work (related literature and data, authors, projects, terms)**
Mark only one oval.

- Yes
- Somewhat
- No

Data completeness and correctness

16. **Granularity of data entities in dataset is appropriate (see p.2 and 3 of Guidelines in Respect of Meta-Data Granularity (see goo.gl/komKJz)**
Mark only one oval.

- Yes
- Somewhat
- No

17. **Structure, size and MIME type of the dataset agrees with description of the dataset content**
Mark only one oval.

- Yes
- Somewhat
- No

18. **Content of the dataset agrees with description of the dataset content**
Mark only one oval.

- Yes
- Somewhat
- No

19. ***Description of methods used to create this dataset are appropriate for the context and discipline**
Mark only one oval.

- Yes
- Somewhat
- No

Findability & interoperability

20. ***Sufficient discovery metadata: metadata includes community accepted keywords and/or terms associated with relevant standards or terminologies**
Mark only one oval.

- Yes
- Somewhat
- No

21. **Dataset is provided in a widely-used or community-accepted machine-readable format and using standard terminologies for nominal data and available standard protocols**
Mark only one oval.

- Yes
- Somewhat
- No