

# 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://goo.gl/kPmYJm>

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:

Currently this checklist is not geared towards all forms of scholarship and would need further generalizing to a wider audience.  
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 dataset.

### 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

- 1. **Email address \***

---

## Dataset and Assessor Identification

- 2. **Repository**

---

- 3. **Location of Dataset (PID or URI preferred)**

\*

---

#### 4. Assessor's Name and e-mail address

---

## Repository Certification

#### 5. Dataset is located within a CoreTrustSeal-certified repository \*

Mark only one oval.

Yes

No

#### 6. Dataset is located within a WDS or Data Seal of Approval certified repository \*

Mark only one oval.

Yes

No

## Curation

#### 7. Structure, harmonization, completeness, and correctness of the dataset comports with level of data curation within this repository as determined by CoreTrustSeal certification (to be verified by evaluator)

Mark only one oval.

Yes

Somewhat

No

## Metadata Completeness

#### 8. Citation exists, including authorship, year, comprehensive title, persistent identifier (e.g. DOI)

Mark only one oval.

Yes

Somewhat

No

9. **\*\*\*Description of the dataset content adequate; this includes the following:**

*Check all that apply.*

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

10. **Coverage (spatial, temporal, or other dimensions) adequate**

*Mark only one oval.*

- Yes
- Somewhat
- No

11. **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)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
a description of data collection or generation (sampling events, used devices, processing steps, QA/QC, experimental or model setup including parameterization)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
References to related work (scholarly work using the data, method papers, references to preceding or related data products)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

12. **\*\*\*Additional metadata adequate to respective research domain (if applicable)**

*Mark only one oval.*

- Yes
- Somewhat
- No

**13. Terms of usage (licenses, other conditions of reuse, data protection, ethical issues)**

*Mark only one oval.*

- Yes  
 Somewhat  
 No

**Accessibility****14. 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****15. Structure, size and MIME type of the dataset agrees with description of the dataset content**

*Mark only one oval.*

- Yes  
 Somewhat  
 No

**16. Content of the dataset agrees with description of the dataset content**

*Mark only one oval.*

- Yes  
 Somewhat  
 No

**17. \*\*\*Validity of methods: methods described to create this dataset are appropriate for the context and discipline**

*Mark only one oval.*

- Yes  
 Somewhat  
 No

18. **Granularity of data entities is appropriate (see p.2 and 3 of Guidelines in Respect of Meta-Data Granularity (see [goo.gl/komKJz](http://goo.gl/komKJz)))**

*Mark only one oval.*

- Yes  
 Somewhat  
 No

## Findability & interoperability

19. **\*\*\*Sufficient discovery metadata: metadata includes community accepted keywords and/or terms associated with relevant standards or terminologies**

*Mark only one oval.*

- Yes  
 Somewhat  
 No

20. **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

---

Powered by

