FAIRsFAIR Data Objects Assessment Metrics
Release Date : 30.04.2020
Version : 0.2

This specification contains **14 metrics** proposed by FAIRsFAIR to evaluate FAIRness of research data in Trustworthy Digital Repositories (TDRs). We developed the metrics based on existing work[1,2,3], and adapted them to accommodate the requirements of two main use cases the project prioritized:

- A TDR will offer a self-assessment tool to educate and raise awareness of researchers on making their data FAIR before depositing the data into the repository.
- A TDR committed to FAIR data provision wants to programmatically evaluate published data for their level of FAIRness over time.

To facilitate both use cases, currently we are developing a manual self-assessment tool and an automated assessment service. The tools will support data FAIRness evaluation based on the metrics proposed. They will be piloted with repositories selected for in-depth collaboration with the project iteratively from 1st May 2020 - 31st August 2021. The metrics are specified following the template below, modified from Wilkinson et al. (2018). The metrics should not be regarded as final but rather as first prioritized criteria for piloting data FAIRness assessment in the context of the use cases above.

Field	Description
Metric Identifier	The local (FAIRsFAIR) identifier[4] of the metric.
Metric Name	The short name of the metric.
Metric Description	The definition of the metric, including its examples and supporting details.
To which FAIR principle(s) does it apply?	The FAIR principle addressed by the metric.
To which CoreTrustSeal requirement(s) does it apply?	The CoreTrustSeal requirements addressed by the metric. One metric may be related to one or more CoreTrustSeal requirements.
For which digital resource is this relevant?	The type of digital resource that will be assessed based on the metric, e.g., data or metadata.

---

[1] RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. Research Data Alliance. DOI: 10.15497/RDA00045

[2] Austin, C., Cousijn, H., Diepenbroek, M., Petters, J., Soares E Silva, M. (2019): WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations. DOI: 10.15497/rda00034

[3] FAIRDat http://blog.ukdataservice.ac.uk/fair-data-assessment-tool/, FAIREnough, https://docs.google.com/forms/d/e/1FAIpQLSf7t1Z9IOBoj5GgWqik8KnhtH3B819Ch6lD5KuAz7yn0I0Opw/viewform

[4] The identifiers should be registered with globally unique identifiers when the metrics are finalized and implemented.

Manual Assessment (User Question)	The question that will be addressed to users as part of the manual-based assessment.
Automated Assessment	The details on the automated assessment of the metric, including inputs, methods and outputs of the assessment.
Comments	A list of related resources, constraints and limitations of the proposed assessment.

1. Globally Unique Identifier

FIELD	DESCRIPTION														
Metric Identifier	FsF-F1-01D														
Metric Name	Globally unique identifier														
Metric Description	A data object may be assigned with a globally unique identifier such that it can be referenced unambiguously by humans or machines. Globally unique means an identifier should be associated with only one resource at any time. Examples of unique identifiers of data are Uniform Resource Locator (URL), Digital Object Identifier (DOI), the Handle System, identifiers.org, w3id.org and Archival Resource Key (ARK). Identifiers are assigned by a data repository (or other service providers) when you make data or metadata available through their services. In return, they will ensure the identifier continues to point to the same data or metadata, according to the specified access terms and conditions.														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
	X														
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
													X			

For which digital resource is this relevant?	Data	Metadata
	X	

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Does the data have a globally unique identifier assigned? ● Yes ● No	Single choice
AUTOMATED ASSESSMENT	

Input	Data identifier
Assessment	Check if the data identifier specified is based on a globally unique identifier scheme.
Output	Assessment status: Yes/No Additional details (if Yes): Identifier scheme

COMMENTS
Resources ● Examples of identifiers compiled by FAIRsharing, https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema ● Uniform Resource Identifier (URI) Schemes, https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml#uri-schemes-1 ● Examples of URI schemes included in rfc3986, https://tools.ietf.org/html/rfc3986#section-1.1.2

2. Persistent Identifier

FIELD	DESCRIPTION														
Metric Identifier	FsF-F1-02D														
Metric Name	Persistent identifier														
Metric Description	We make a distinction between the uniqueness and persistence over time of an identifier. An HTTP URL (the address of a given unique resource on the web) is globally unique, but is not persistent, whereas a DOI is both globally unique and persistent. A persistent identifier should be maintained and governed such that it remains stable and resolvable for the long term. For example, the DOI system guarantees the persistence of its identifiers through its social (policy) and technical infrastructures. The persistent identifier of a data object may be resolved to a data file, a web service response that contains data values, or to a proxy (e.g., an online page that contains metadata, including the link to access the actual data).														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
	X														
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
													X			
For which digital resource is this relevant?	Data								Metadata						
	X														
MANUAL ASSESSMENT (USER QUESTION)								QUESTION TYPE							

Does the data have a persistent identifier assigned? ● Yes ● No	Single choice

AUTOMATED ASSESSMENT	
Input	Data identifier
Assessment	Check if the data identifier specified is based on a commonly accepted persistent identifier scheme, and it resolves to a web address where the data can be found. Test the identifier resolvability through the HTTP response status codes.
Output	Assessment status: Yes/No Additional details (if Yes): Persistent identifier scheme, identifier resolvable status and resolved URL.

COMMENTS

Resources
● A wiki entry on persistent identifier, https://en.wikipedia.org/wiki/Persistent_identifier
● Generic PID definitions, Initial Persistent Identifier Policy for the EOSC, https://doi.org/10.5281/zenodo.3574202
● FREYA Deliverable 3.1 (Survey of Current PID Services Landscape), https://doi.org/10.5281/zenodo.1324295

Known Limitations/Constraints
● The automated assessment verifies the resolvability of the specified identifier, but does not assert the type of the resolved object, e.g., whether the resolved object is a landing page, a data file or a web service response.
● The persistence policy of a PID is identified manually before the PID is included in the list of commonly accepted persistent identifier schemes that will be used by the automated assessment.

3. Descriptive Core Metadata

FIELD	DESCRIPTION													
Metric Identifier	FsF-F2-01M													
Metric Name	Descriptive core metadata													
Metric Description	Metadata is descriptive information about a data object. Since the metadata required depends on users and their applications, this metric focuses on core metadata, which is the minimum descriptive information required to enable data citation and discovery. We determine the required metadata based on common data citation guidelines, (e.g., DataCite, ESIP, and IASSIST), and metadata recommendations for data discovery, (e.g., EOSC Datasets Minimum Information (EDMI), DataCite Metadata Schema, and W3C Recommendation Data on the Web Best Practices). This metric focuses on domain-agnostic core metadata. Domain or discipline-specific metadata specifications are covered under metric FsF-R1.3-01M.													
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
		X													

To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
													X			

For which digital resource is this relevant?	Data			Metadata		
				X		

MANUAL ASSESSMENT (USER QUESTION)	**QUESTION TYPE**
Are metadata elements provided to support data citation and discovery (creator, title, data identifier, publisher, publication date, summary and keywords describing the data)? ● Not provided ● Partially provided ● Completely provided	Single choice

AUTOMATED ASSESSMENT	
Input	Data identifier
Assessment	Parse or retrieve metadata, e.g., through the options below, and then verify presence/absence of the core elements in the metadata. ● Structured data embedded in the landing page of the identifier (e.g., Schema.org, Dublin Core and OpenGraph meta tags) ● Typed Links in the HTTP Link header; for more information, see https://signposting.org/conventions/ ● Use the identifier to retrieve its metadata from the DataCite Search using different standards; see DataCite Content Resolver at https://datacite.org/content.html.
Output	Assessment status: no metadata, partial metadata or all metadata Additional details (if metadata found): The key-value pairs of core metadata elements and their values, and the sources of the metadata (e.g., Schema.org)

COMMENTS

Resources
● Examples of metadata recommendations:
  ○ EOSC EDMI metadata properties, https://eosc-edmi.github.io/properties
  ○ W3C Recommendation Data on the Web Best Practices, https://www.w3.org/TR/dwbp/#metadata
● Sites that provide a list of metadata standards:
  ○ RDA Metadata Standards Catalog, https://rdamsc.bath.ac.uk/
  ○ FAIRsharing standards, https://fairsharing.org/standards/
  ○ DCC List of Metadata Standards, http://www.dcc.ac.uk/resources/metadata-standards/list
● Examples of domain agnostic metadata standards for describing research data:
  ○ Dublin Core Metadata Initiative (DCMI) Metadata Terms, https://www.w3.org/TR/dwbp/#bib-DCTERMS
  ○ DataCite Metadata Schema, https://doi.org/10.14454/7xq3-zf69
  ○ Schema.org, https://schema.org/Dataset

- ○ Data Catalog Vocabulary (DCAT), https://www.w3.org/TR/dwbp/#bib-VOCAB-DCAT

Known Limitations/Constraints
- The automated assessment assumes that the identifier resolves to a landing page (e.g., html) that contains the metadata of the data. Landing page may not necessarily be an html page.
- Data providers may use different standards to expose the metadata of their data.
- The metadata records maintained by a data provider might not be accessible, due to, e.g., broken link of the landing page, proprietary metadata standard used, and restricted metadata).

4. **Inclusion of Data Identifier in Metadata**

FIELD	DESCRIPTION														
Metric Identifier	FsF-F3-01M														
Metric Name	Inclusion of data identifier in metadata														
Metric Description	The metadata should include the identifier of the data such that users can discover and access the data through the metadata.														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
			X												
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
													X			
For which digital resource is this relevant?	Data								Metadata							
									X							

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Does the metadata include the data identifier? ● Yes ● No	Single choice

AUTOMATED ASSESSMENT	
Input	Data Identifier
Assessment	Verify if the identifier to access the actual data is included through selected elements in the metadata (e.g., 'Distribution' property of Schema.org), or through the Typed Links with, for instance

	'foaf:isPrimaryTopicOf', 'describes', 'item' relation type. Check if the identifier is 'active' through the HTTP response status codes.
Output	Assessment status: Yes/No Additional details (if Yes): Identifier included and its status (i.e., active or broken link).

COMMENTS

Resources
● Relation Types for Typed Links, https://signposting.org/conventions/
 	
Known Limitations/Constraints
● A metadata standard may not support any element or include multiple elements through which a data identifier may be specified.
● Different practices of associating data with its metadata should be handled as part of the automated assessment:
    ○ Data is assigned with an identifier that resolves to a page that contains metadata of the data. The metadata may contain the identifier and a URL to access the data (contents). In this case, the access URL should be tested.
    ○ Data and metadata are assigned with separate identifiers. Therefore, the data identifier should be tested.

5. Searchable Metadata

FIELD	DESCRIPTION														
Metric Identifier	FsF-F4-01M														
Metric Name	Searchable metadata														
Metric Description	This metric refers to various ways through which the metadata of data is exposed or offered in a standard and machine-readable format. Answering this metric will require an understanding of the capabilities offered by the data repository used to host the data. For example, if data is hosted by a generic or domain/discipline specific repository, the repository may expose its metadata to allow harvesting through a specific protocol (e.g., via OAI-PMH) and/or a web service. Metadata may also be embedded as structured data (e.g., schema.org implementation) on a data page for use by web search engines such as Google and Bing. You can, for instance, use the Google Dataset Search (https://datasetsearch.research.google.com/) to check if data hosted by the repository of your choice are indexed.														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
				X											
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
													X			

For which digital resource is this relevant?	Data	Metadata
		X

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Is the metadata offered in such a way that it can be harvested? ● Metadata is not offered ● Metadata is offered through a web service or machine harvestable ● Metadata is offered as structured data on the data page for use by a web search engine	Multiple choice (2,3 options)

AUTOMATED ASSESSMENT	
Input	Data Identifier, metadata access endpoint (if it is not included in the metadata or landing page of the identifier).
Assessment	Check if metadata access endpoint returns metadata records based on a request using the data identifier. Check if search engine friendly structured data is embedded in the data landing page with a proper resource type, e.g., schema.org representation of type 'Dataset' or 'Collection').
Output	Assessment status: Yes/No Additional details (if Yes): Mechanisms through which metadata is offered (e.g., web service, structured data)

COMMENTS

Resources
● Google reference documentation on representing structured data of Dataset, https://developers.google.com/search/docs/data-types/dataset

Known Limitations/Constraints
● Data providers may expose their metadata through different ways, e.g., OAI-PMH, REST API using JSONAPI specification, and Catalog Service for the Web (CSW). Their endpoints (URLs) should be machine discoverable and accessible. The metadata access endpoints of a repository can be found through FAIRsharing and re3data. However, at present, it is not possible to programmatically discover the metadata endpoints of a repository based on a data identifier, unless they are explicitly specified in the metadata or the landing page of the data. Mapping DataCite repositories (formerly clients) to re3data identifiers in progress.
● Structured data may be represented in different formats, JSON-LD, Microdata, and RDFa. The variety of formats should be handled as part of the automated assessment.
● The automated assessment only verifies if structured data is present on the data landing page with a proper type (e.g., Dataset or Collection). Embedding structured data does not guarantee that the data will be present on search results. To verify that the data is findable through a web search engine, we should perform a search through the search engine API based on the data identifier and its descriptive metadata (e.g., title, author). However, most of the web search engine APIs (e.g., Google Custom Search, Bing Web Search API) offers a limited number of free search queries.

6. Data Access Level

FIELD	DESCRIPTION
Metric Identifier	FsF-A1-01M
Metric Name	Data access level
Metric Description	This metric determines if the metadata includes the level of access to the data such as public, embargoed, restricted, or closed access. It is recommended that data should be as open as possible and as closed as necessary. ● Datasets should be public domain and openly accessible without restrictions when possible. ● Embargoed access refers to data that will be made publicly accessible at a specific date which should be specified in the metadata. For example, a data author may release their data after having published their findings from the data. ● Restricted access refers to data that can be accessed under certain conditions (e.g. because of commercial, sensitive, or other confidentiality reasons or the data is only accessible via a subscription or a fee). Restricted data may be available to a particular group of users or after permission is granted. For restricted data, the metadata should include the conditions of access to the data (e.g., point of contact or instructions to access the data). ● Closed access refers to data that is not made publicly available and for which only metadata is publicly available.

To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
					X										

To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
															X	

For which digital resource is this relevant?	Data	Metadata
		X

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Select the access level of the data and specify if the metadata includes the access level, and its related access information (if required). ● Public access ● Embargoed access 　If the data is embargoed, does the metadata include the date the data will be released publically? 　▪ Yes 　▪ No ● Restricted access	Single choice

	If access to the data is restricted, does the metadata include access conditions, e.g., point of contact or instructions to access the data? ▪ Yes ▪ No ● Closed access (metadata only) ● No option to select an access level	

AUTOMATED ASSESSMENT		
Input	Data Identifier, metadata access endpoint (if it is not included in the metadata or landing page of the identifier).	
Assessment	Check the presence/absence of data access level through metadata element(s). If it is embargoed data, check if the embargo end date is specified. If it is restricted data, check if the data access conditions are specified.	
Output	Assessment status: Yes/No (this depends on the access level and its related access information) Additional details (if Yes): Access level found (public, embargoed, restricted, metadata only, none), and data access information included in the metadata (if applicable).	

COMMENTS		

Resources
- Creative Commons License (may indicate 'open access'), https://creativecommons.org/share-your-work/licensing-examples/
- EU Vocabulary on access rights, https://op.europa.eu/en/web/eu-vocabularies/at-dataset/-/resource/dataset/access-right
- Open Digital Rights Language (ODRL) Information Model 2.2, https://www.w3.org/TR/odrl-model/
- Controlled Vocabulary for Access Rights, http://vocabularies.coar-repositories.org/documentation/access_rights/
- Archival Access Rights Vocabulary (test vocabulary, not yet available through the production metadata registry), http://sandbox.metadataregistry.org/concept/list/vocabulary_id/251.html
- Eprints Access Rights Vocabulary Encoding Scheme, http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_AccessRights_Vocabulary_Encoding_Scheme

Known Limitations/Constraints
- The metadata standard considered as part of the assessment may not include all of the elements for representing data access levels and related access information.
- The access information may be expressed in an unstructured manner, e.g., as a 'comment' in the metadata document.
- This assessment should be complemented with the evaluation of the data access mechanism based on the specified access levels, e.g., data is not accessible, accessible in a semi-automated (mediated access to data via data custodian), or automated fashion.

7. Metadata Preservation

FIELD	DESCRIPTION
Metric Identifier	FsF-A2-01M

Metric Name	Metadata preservation
Metric Description	This metric determines if the metadata will be preserved even when the data they represent are no longer available or lost. Similar to metric FsF-F4-01M, answering this metric will require an understanding of the capabilities offered, data management plan (DMP) and policies implemented by the data repository. Continued access to metadata depends on a data repository's preservation practice which is usually documented in the repository's service policies or statements. Typically, a trustworthy data repository offering DOIs and implementing a PID Policy will guarantee that metadata will remain accessible even when data is no longer available for any reason (e.g., by providing a tombstone page)

To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
								X							

To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
										X						

For which digital resource is this relevant?	Data		Metadata	
			X	

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Will you deposit your data in a repository which ensures that the metadata remains available once the data becomes unavailable? ● Yes ● No	Single choice

AUTOMATED ASSESSMENT

Input	-
Assessment	Not applicable, see comments below.
Output	-

COMMENTS

Resources
● DMPonline, https://dmponline.dcc.ac.uk/public_plans
● DMP Common Standards WG, https://www.rd-alliance.org/groups/dmp-common-standards-wg
● ezDMP, https://ezdmp.org/index
● Best Practices for offering tombstone pages, https://support.datacite.org/docs/tombstone-pages

Known Limitations/Constraints
● Programmatic assessment of the preservation metadata of a dataset can only be tested when it is deleted or replaced. So this test is only applicable for deleted or obsolete datasets. Importantly, continued access to metadata depends on a data repository's preservation practice. Therefore, we regard that the assessment of metric applies to at the level of a repository, not at the level of individual objects. For this reason, we excluded its automated assessment from this

specification. Nonetheless, to raise awareness of users about their choice of a repository, we include the metric as part of the manual assessment.
- Data preservation statements are usually found in a repository's data policy or other governance documents. Machine-actionability of these documents is important to enable an automated assessment of the statements. Several groups/projects have initiated the idea (e.g., DMP Common Standards WG, ezDMP, DCC's DMPRoadmap project) but more effort is required to translate it into practice.
- Currently, PID providers (e.g., DataCite) do not offer any tombstone pages automatically for deleted objects. Data providers may maintain the pages instead, for example https://doi.pangaea.de/10.1594/PANGAEA.715333

8. Semantic Representation of Metadata

FIELD	DESCRIPTION															
Metric Identifier	FsF-I1-01M															
Metric Name	Semantic representation of metadata															
Metric Description	To make metadata more understandable to humans and machines, they are described with semantic vocabularies. Ontology, thesaurus, and taxonomy are kinds of semantic vocabularies, and they come with different degrees of expressiveness, structure, and inferential power. Metadata may use semantic vocabularies in various ways. For example, semantic vocabularies may be embedded in the metadata page (e.g., Microdata, RDFa or JSON-LD). Metadata may also be published as Linked Data using semantic vocabularies.															
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3	
									X							
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
															X	
For which digital resource is this relevant?	Data								Metadata							
									X							

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Are semantic vocabularies (e.g., ontologies, thesauri, taxonomies) used in the metadata ? ● No ● Yes, semantic vocabularies embedded in the metadata page ● Yes, metadata is available as Linked Data	Multiple choice (2,3 options)

AUTOMATED ASSESSMENT	
Input	Data Identifier, SPARQL endpoint (if supported)

Assessment	Test if data landing page is semantically annotated, e.g., using RDFa, Microdata, JSON-LD Test if metadata of the data is available as Linked Data (include accept header with RDF content types).
Output	Assessment status: Yes/No Additional details (if Yes): Vocabulary application status (embedded in metadata page, linked data)

COMMENTS

Resources
- A list of content types is available at https://www.iana.org/assignments/media-types/media-types.xhtml
- SPARQL Protocol for RDF, https://www.w3.org/TR/rdf-sparql-protocol/

Known Limitations/Constraints
- The automated assessment checks the inclusion of semantic markup in the metadata page, not their contents, e.g., if the vocabularies used are in appropriate context and accessible over the web.
- RDF data may be expressed in a number of different ways, e.g., RDF/XML, turtle, n3 (extension of turtle), ntriples, and JSON. Therefore, the variety of serialization formats (and their respective MIME types) should be considered as part of the automated assessment.

9. Meaningful Links to Related Entities

FIELD	DESCRIPTION														
Metric Identifier	FsF-I3-01M														
Metric Name	Meaningful links to related entities														
Metric Description	Linking data to its related entities will increase its potential for reuse. The linking information should be captured as part of the metadata. A data may be linked to its prior version, other relevant datasets, related publications, source (such as an instrument), data creators or collectors and organization (e.g., funder and hosting institution). Links between data and its related entities should be meaningful and expressed through relation types (e.g., the DataCite Metadata Schema specifies relation types between research objects through the fields 'RelatedIdentifier' and 'RelationType').														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
											X				
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
											X					
For which digital resource is this relevant?	Data				Metadata										
					X										

13

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Does the metadata include meaningful links between the data and its related entities? ● Yes ● No	Single choice

AUTOMATED ASSESSMENT	
Input	Data Identifier
Assessment	Check the metadata elements which indicate the relationship between data and related entities. Test if the URLs of the related entities are active (not broken links).
Output	Assessment status: Yes/No Additional details (if Yes): The related entities, and the relationships specified.

COMMENTS
Resources ● The DataCite Metadata Schema specifies relation types between research entities, https://schema.datacite.org/meta/kernel-4.3/include/datacite-relationType-v4.xsd ● Link Relation Types, https://www.iana.org/assignments/link-relations/link-relations.xhtml **Known Limitations/Constraints** ● Different metadata schemas may use different properties to specify the relation between data and its related entities. ● The automated assessment regards any relation between a data and its related entities as success. It does not consider the quantity or types of relations.

10. Data Content Description

FIELD	DESCRIPTION														
Metric Identifier	FsF-R1-01MD														
Metric Name	Data content description														
Metric Description	This metric evaluates if the content descriptions are specified in the metadata and if they conform with the actual data. Examples of content descriptions are resource type (e.g., dataset or collection), variable(s) measured, method, data format and size.														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
												X			
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
											X					

For which digital resource is this relevant?	Data	Metadata
	X	X

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Are content descriptions specified in the metadata? ● Yes ● No If yes, do they conform with the data? ● Fully conforming ● Somewhat conforming ● Not conforming	Single choice

AUTOMATED ASSESSMENT	
Input	Data identifier
Assessment	Verify the presence/absence of elements representing content descriptions in the metadata. Use the data access URL specified in the metadata to retrieve the actual data. Compare the content descriptions found with data properties (see comments below).
Output	Assessment status: Yes/No Additional details (if Yes): Content descriptions specified, and their conformance status (full, partial, none).

COMMENTS

Resources
● Model for Tabular Data and Metadata on the Web, https://www.w3.org/TR/tabular-data-model

Known Limitations/Constraints
● General-purpose metadata standards such as Datacite Metadata Schema and Schema.org provide elements to represent content descriptions. Thus, it is possible to check programmatically if the descriptions required are present in the metadata. However, the conformance test may become a challenge due to a variety of data types and data size. Standardized tabular data and self-describing data formats (e.g., HDF, NetCDF, Parquet) are promising, but not the solution to every research domain.
● At present, the proposed automated assessment is limited to the evaluation of the size and format of the data. Content descriptions (e.g., method, variable measured) cannot be assessed programmatically due to the challenge of parsing different types of data, and unstructured content descriptions might be included in a data file (fuzzy text-matching algorithms can be useful here).

11. **Data Usage License**

FIELD	DESCRIPTION
Metric Identifier	FsF-R1.1-01M
Metric Name	Data Usage License
Metric Description	This metric evaluates if data is licensed because otherwise users cannot reuse it in a clear legal context. We encourage the application of licenses for all kinds of data whether public, restricted or for specific users. Without an explicit license, users do not have a clear idea of what can be done with your data. Licenses can be of standard type (e.g., Creative Commons) or bespoke licenses, and rights statements which indicate the conditions under which data can be reused. It is highly recommended to use a standard, machine-readable license such that it can be interpreted by machines and humans. In order to inform users about what rights they have to use a dataset, the license information should be specified as part of the dataset's metadata.

To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
													X		

To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
		X														

For which digital resource is this relevant?	Data	Metadata
		X

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Does the metadata include license information under which the data can be reused? ● Yes ● No	Single choice

AUTOMATED ASSESSMENT	
Input	Data Identifier
Assessment	Verify the presence/absence of metadata element(s) corresponding to license information. Use the license information (e.g., name or URI), to request additional information from the SPDX license registry.
Output	Assessment status: Yes/No Additional details (if Yes): License specified, and additional information retrieved from the license registry, if available.
COMMENTS	

Resources
- Common licenses are available at SPDX license registry, https://spdx.org/licenses/
- Rights statements of cultural heritage objects, https://rightsstatements.org/page/1.0/?language=en
- ARDC Data Rights Management Guide, https://ardc.edu.au/guides/research-data-rights-management
- The Landscape of Rights and Licensing Initiatives for Data Sharing, https://doi.org/10.5334/dsj-2019-029

Known Limitations/Constraints
- The automated assessment checks if the license information is provided as part of the metadata. It does not validate if the specified license is the most appropriate license for the data.

12. Data Provenance

FIELD	DESCRIPTION														
Metric Identifier	FsF-R1.2-01M														
Metric Name	Data provenance														
Metric Description	It is essential to provide provenance information of your data to enable its use and reuse. Data provenance (also known as lineage) represents its history, including people, entities, and processes involved in the data creation. It depends on the data type (e.g., measurement, observation, derived data, or data product) and research domains. For that reason, it is difficult to capture a set of finite provenance records adequate to all domains. Based on existing work, we suggest that the following provenance properties of data generation or collection should be supplied as part of the metadata (this is not an exhaustive list). ● Sources of data generation or collection, e.g., model, instrument, etc. ● Data creation or collection date ● Contributor involved ● Data versioning information														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
														X	
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
							X									

For which digital resource is this relevant?	Data		Metadata	
			X	

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Is provenance information on collection or generation of data (sources, date, contributor, version) included in the metadata? ● Not provided	Single choice

17

● Partially provided ● Completely provided	

AUTOMATED ASSESSMENT	
Input	Data Identifier
Assessment	Verify the presence/absence of metadata element(s) corresponding to data provenance properties.
Output	Assessment status: no metadata, partial metadata or all metadata Additional details (if Yes): The key-value pairs of metadata elements representing data provenance and their values.

COMMENTS

Resources
- PROV Model Primer, https://www.w3.org/TR/prov-primer/
- Checklist for Evaluation of Dataset Fitness for Use produced by the WDS/RDA Assessment of Data Fitness for Use WG, https://www.rd-alliance.org/system/files/DataFitnessForUse_ChecklistForm_v2_20181218_RDADistribution.pdf
- W3C Recommendation Data on the Web Best Practices (8.4 Data Provenance), https://www.w3.org/TR/dwbp/#metadata
- DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. https://doi.org/10.14454/7xq3-zf69

Known Limitations/Constraints
- The proposed provenance properties are not final; new properties may be incorporated into the assessment if the requirement emerges.
- We regard references to related work (scholarly articles, data papers, preceding or associated data) as useful provenance information. This property of provenance is considered as part of FsF-I3-01M, therefore we excluded it from the assessment.
- Metadata may include a specific element (e.g., dcmi:provenance) and/or 'proxy' elements (e.g., datacite:Contributor, schema.org:measurementTechnique) to convey data provenance.

13. Community-Driven Metadata

FIELD	DESCRIPTION
Metric Identifier	FsF-R1.3-01M
Metric Name	Community-driven metadata
Metric Description	In addition to core metadata required to support data citation and discovery covered under metric FsF-F2-01M, metadata to support data reusability should be made available following community-endorsed metadata standards. Community metadata standards may exhibit different levels of readiness. Some communities have well-established metadata standards (e.g., geospatial: ISO19115, biodiversity: DarwinCore, ABCD, EML, social science: DDI, astronomy: International Virtual Observatory Alliance Technical Specifications). In contrast, others, including new domains, may have limited standards or standards that are under development (e.g., engineering and linguistics).

	The use of community-endorsed metadata standards is usually encouraged and supported by domain and discipline-specific repositories.														
To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
															X
To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
														X		
For which digital resource is this relevant?	**Data**						**Metadata**								
							X								

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Does the metadata follow the specifications of a community-endorsed standard? ● Yes ● No ● Standard unknown/unavailable	Single choice

AUTOMATED ASSESSMENT	
Input	Data Identifier, Metadata access endpoint
Assessment	Gather all metadata standards used by a data repository for disseminating metadata; this list can be requested, e.g., from the metadata endpoint (e.g., OAI-PMH). Filter out domain-agnostic standards (e.g., Datacite Metadata Schema, Dublin Core, Schema.org) from the list. Request metadata of the data identifier specified based on one (test case) of the remaining standards.
Output	Assessment status: Yes/No Additional details (if Yes): Prefix and namespace of all metadata standards supported by the repository, test case request status.

COMMENTS

Resources
Examples of the metadata standards with subject areas:
● RDA Metadata Standards Catalog, https://rdamsc.bath.ac.uk/
● FAIRSharing, https://fairsharing.org/standards/

Known Limitations/Constraints
● The automated assessment focuses on a specific metadata harvesting protocol. It might not be supported by all data repositories.
● The assessment should be extended to identify the subject area(s) of the metadata standards and provide the information as part of the assessment output.
● Future evaluation of the metric should also consider metadata completeness, i.e., the degree to which the metadata is specified based on a community-endorsed standard.

14. Data File Format

FIELD	DESCRIPTION
Metric Identifier	FsF‑R1.3‑02D
Metric Name	Data File format
Metric Description	File formats refer to methods for encoding digital information. For example, CSV for tabular data, NetCDF for multidimensional data and GeoTIFF for raster imagery. Data should be made available in a preferred file format that is accepted by the research community to enable data sharing and reuse. Preferred formats are formats that are widely used and supported by the most commonly used software and tools. Preferred formats not only give a higher certainty that your data can be read in the future, but they will also help to increase the reusability and interoperability. Using preferred formats enables data to be loaded directly into the software and tools used for data analysis. It makes it possible to easily integrate your data with other data using the same preferred format. The use of preferred formats will also help to transform the format to a newer one, in case a preferred format gets outdated.

To which FAIR principle(s) does it apply?	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
															X

To which CoreTrustSeal requirement(s) does it apply?	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
															X	

For which digital resource is this relevant?	Data	Metadata
	X	

MANUAL ASSESSMENT (USER QUESTION)	QUESTION TYPE
Is the data available in a preferred file format? ● Yes ● No ● Format unknown	Single choice

AUTOMATED ASSESSMENT	
Input	Data Identifier
Assessment	Extract file format information from the metadata based on elements, e.g., datacite:Format, schema.org: fileFormat, dc:format. Check if the format is a preferred format (see comments below).
Output	Assessment status: Yes/No Additional details (if Yes): File format(s) specified in the metadata

COMMENTS

Resources
- Examples of recommended file formats based on data types,
  https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx
- PRONOM file format registry,
  https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new
- List of open formats, https://en.wikipedia.org/wiki/List_of_open_formats
- A list of common media types, https://www.iana.org/assignments/media-types/media-types.xhtml

Known Limitations/Constraints
- At present, preferred file formats by communities are not available through a registry but on static web pages (see resources above). The resources are incomplete. Further work is necessary to gather feedback from communities on their preferred file formats and then incorporate this information into the assessment.