

Testing the FAIR metrics on data catalogs

Supplementary File: FAIR metrics reviews

FAIR metrics review Dataverse

Executed by Jarno van Erp, starting 29 December 2017, using <https://dataverse.harvard.edu>

Metric identifier	What is being measured?	What must be provided?	Score Manual	Score Automated
https://purl.org/fair-metrics/FM_F1A (Identifier Uniqueness)	Whether there is a scheme to uniquely identify the digital resource.	URL to a registered identifier scheme. <i>URL:</i> (http://dublincore.org/documents/dcmi-terms/)	Dataverse uses Dublin core for identifying. This is registered in the metadata. Present	Present integrated within Dataverse
https://purl.org/fair-metrics/FM_F1B (Identifier persistence)	Whether there is a policy that describes what the provider will do in the event an identifier scheme becomes deprecated.	A URL that resolves to a document containing the relevant policy.	Not clear if they mean that Dataverse should specify this or if the identifier scheme provider should specify this. For Dataverse it is absent .	Absent
https://purl.org/fair-metrics/FM_F2 (Machine-readability of metadata)	The availability of machine-readable metadata that describes a digital resource.	A URL to a document that contains machine-readable metadata for the digital resource.	Can be done however very dependent on uploader. For software I would consider this as Present	Is automatically generated from the metadata. Present

		<p>Furthermore, the file format must be specified.</p> <p>URL: https://dataversen.nl/api/dataset/export?exporter=dcterms&persistentId=hdl%3A10411/UQZGXY</p> <p>Format specification:</p> <pre><metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns="http://dublincore.org/documents/dcmi-terms/"></pre> <p>In the heading of metadata</p>		
https://purl.org/fair-metrics/FM-E3 (Resource Identifier in Metadata)	<p>Whether the metadata document contains the globally unique and persistent identifier for the digital resource.</p>	<p>The URL of the metadata and the GUID of the digital resource it describes.</p> <p>URL to metadata: https://dataversen.nl/api/dataset/export?exporter=dcterms&persistentId=hdl%3A10411/UQZGXY</p>	<p>Resource is identified</p> <p>Present</p>	<p>Automatically added to Dublin core.</p> <p>Present</p>

		<p>e.nl/api/dataset/export?exporter=dcterms&persistentId=hdl%3A10411/UQZGXY</p> <p>GUID:</p> <p><dcterms:identifier>hdl:10411/UQZGXY</dcterms:identifier></p>		
<p>https://purl.org/fair-metrics/FME4 (Indexed in a searchable resource)</p>	<p>The degree to which the digital resource can be found using web-based search engines</p>	<p>The persistent identifier of the resource and one or more URLs that give search results of different search engines.</p> <p>Persistent identifier:</p> <p>hdl:10411/UQZGXY</p> <p>URL:</p> <p>https://duckduckgo.com/?q=hdl%3A10411%2FUQZGXY&t=h&ia=web</p>	<p>Resource can be found through a web-based search engine. However you need to keep indexing in mind when deploying a Dataverse instance.</p> <p>Present</p>	<p>Present</p>
<p>https://purl.org/fair-metrics/FMA1.1 (Access Protocol)</p>	<p>The nature and use limitations of the access protocol.</p>	<p>i) A URL to the description of the protocol</p> <p>Not provided but uses https</p> <p>ii) true/false as to whether the protocol is open source</p> <p>True</p> <p>iii) true/false as</p>	<p>No URL provided to describe HTTPS. I think this is especially needed when the protocol is not common.</p>	<p>Absent</p>

		to whether the protocol is (royalty) free True		
https://purl.org/fair-metrics/FMA1.2 (Access authorization)	Specification of a protocol to access restricted content.	i) true/false concerning whether authorization is needed True ii) a description of the process to obtain access to restricted content. False	Present that authorization is needed. Absent how to get it	Present response no access. Absent machine-readable way how to get access
https://purl.org/fair-metrics/FMA2 (Metadata Longevity)	The existence of metadata even in the absence/removal of data	URL to a formal metadata longevity plan	Metadata is deleted with data Absent	Absent
https://purl.org/fair-metrics/FM1 (Use a Knowledge Representation Language)	Use of a formal, accessible, shared, and broadly applicable language for knowledge representation.	URL to the specification of the language URL: http://dublincore.org/documents/dcmi-terms/ http://purl.org/dc/elements/1.1 http://purl.org/dc/terms/	Present	Automatically included present
https://purl.org/fair-metrics/FM1	The metadata values and qualified	UUIDs representing the vocabularies	For metadata present for data itself	Metadata: Present

<p>2 (Use FAIR Vocabularies)</p>	<p>relations should themselves be FAIR, for example, terms from open, community-accepted vocabularies published in an appropriate knowledge-exchange format.</p>	<p>used for (meta)data</p> <p>xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"</p>	<p>determined by the owner.</p>	<p>Data: Absent</p>
<p>https://purl.org/fair-metrics/FM_13 (Use Qualified References)</p>	<p>Relationships within (meta)data, and between local and third-party data, have explicit and 'useful' semantic meaning</p>	<p>Linksets (in the formal sense) representing part or all of your resource</p>	<p>Possibility to make it present for some relations but absent to the extend/freedom needed.</p>	<p>Present for the once you specified</p>
<p>https://purl.org/fair-metrics/FM_R1.1 (Accessible Usage License)</p>	<p>The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents</p>	<p>GUID of the license (e.g. its URL) for the data license and for the metadata license</p>	<p>Absent only the name of the license is provided not a GUID.</p>	<p>Absent</p>
<p>https://purl.org/fair-metrics/FM_R1.2 (Detailed Provenance)</p>	<p>That there is provenance information associated with the data, covering at least two primary types of provenance information:</p> <ul style="list-style-type: none"> - Who/what/Whe 	<p>Two URLs (GUIDs). One of these URLs points to one of the vocabularies used to describe citational provenance (e.g. dublin core). The second points</p>	<p>Present but it requires additional input from the user.</p>	<p>Absent</p>

	<p>n produced the data (i.e. for citation)</p> <p>- Why/How was the data produced (i.e. to understand context and relevance of the data)</p>	<p>to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM)</p> <ul style="list-style-type: none"> - Dublin core for citational - Possible to provide contextual provenance 		
https://purl.org/fair-metrics/EM_R1.3 (Meets Community Standards)	<p>Certification, from a recognized body, of the resource meeting community standards.</p>	<p>A certification saying that the resource is compliant</p>	<p>Absent</p>	<p>Absent</p>

Overall score FAIRness Dataverse:

Principle:	Score manual:	Score automatic:	Missing:
Findable	4.5/5	4.5/5	F1B
Accessible	1/3(?)	1/3 (?)	A1.1, A2

Interoperable	2.5/3	2/3	I3
Re-usable	1/3	0/3	R1.1, R1.3

Conclusion FAIR metrics review Dataverse:

High findability and possibility to become interoperable when the user puts in the effort.
Re-usability for a user is high but not or a machine.

Main concerns:

- Data license not machine readable
- Certification for community standards needs to be implemented
- Metadata not persisting

FAIR metrics review CKAN

Executed by Jarno van Erp, starting 29 December 2017, using
<https://demo.ckan.org/dataset/newcastle-city-council-payments-over-500>

Metric identifier	What is being measured?	What must be provided?	Score Manual	Score Automated
https://purl.org/fair-metrics/FM_E1A (Identifier Uniqueness)	Whether there is a scheme to uniquely identify the digital resource.	URL to a registered identifier scheme.	Present. CKAN lets users upload documents. Therefore the user himself can upload their RDF DCAT/turtle file. Also the ckanext-dcat plugin exposes the data through RDF.	Absent. Plugin needed
https://purl.org/fair-metrics/FM_	Whether there is a policy that	A URL that resolves to a	Not clear if they mean that	Not clear if they mean that CKAN

<p>F1B (Identifier persistence)</p>	<p>describes what the provider will do in the event an identifier scheme becomes deprecated.</p>	<p>document containing the relevant policy.</p>	<p>CKAN should specify this or if the identifier scheme provider should specify this. For CKAN it is absent.</p>	<p>should specify this or if the identifier scheme provider should specify this. For CKAN it is absent.</p>
<p>https://purl.org/fair-metrics/FM_F2 (Machine-readability of metadata)</p>	<p>The availability of machine-readable metadata that describes a digital resource.</p>	<p>A URL to a document that contains machine-readable metadata for the digital resource. Furthermore, the file format must be specified.</p> <p>URL:</p> <p>https://demo.ckan.org/dataset/newcastle-city-council-payments-over-500.ttl</p> <p>Format specification:</p> <p>"@prefix dcat: <http://www.w3.org/ns/dcat#> ."</p> <p>(and more)</p> <p>In the heading of metadata</p>	<p>Can be done however very dependent on uploader. For software I would consider this as Present</p>	<p>Is automatically generated from the metadata but not for data</p>
<p>https://purl.org/fair-metrics/FM_F3 (Resource Identifier in Metadata)</p>	<p>Whether the metadata document contains the globally unique and persistent identifier for the digital resource.</p>	<p>The URL of the metadata and the GUID of the digital resource it describes.</p> <p>URL to</p>	<p>Resource is identified</p> <p>Present</p>	<p>Resource is identified</p> <p>Present</p>

		<p>metadata:</p> <p>https://demo.ckan.org/dataset/newcastle-city-council-payments-over-500.ttl</p> <p>GUID:</p> <p>dct:identifier "b9076e40-80ea-480b-b330-e399d7a8c09b" ;</p>		
<p>https://purl.org/fair-metrics/FM_F4 (Indexed in a searchable resource)</p>	<p>The degree to which the digital resource can be found using web-based search engines</p>	<p>The persistent identifier of the resource and one or more URLs that give search results of different search engines.</p> <p>Persistent identifier:</p> <p>b9076e40-80ea-480b-b330-e399d7a8c09b</p> <p>URL:</p> <p>https://www.google.nl/search?q=b9076e40-80ea-480b-b330-e399d7a8c09b&oeq=b9076e40-80ea-480b-b330-e399d7a8c09b&aqs=chrome..69j57j69i60.664j0j1&sourceid=chrome&ie=UTF-8</p>	<p>Resource can be found through a web-based search engine.</p> <p>Present</p>	<p>Resource can be found through a web-based search engine.</p> <p>Present</p>
<p>https://purl.org/fair-metrics/FM_A1.1 (Access Protocol)</p>	<p>The nature and use limitations of the access protocol.</p>	<p>i) A URL to the description of the protocol</p> <p>Not provided</p>	<p>No URL provided to describe HTTPS. I think this is especially</p>	<p>No URL provided to describe HTTPS. I think this is especially needed when the</p>

		<p>but uses https</p> <p>ii) true/false as to whether the protocol is open source</p> <p>True</p> <p>iii) true/false as to whether the protocol is (royalty) free</p> <p>True</p>	needed when the protocol is not common.	protocol is not common.
https://purl.org/fair-metrics/FMA1.2 (Access authorization)	Specification of a protocol to access restricted content.	<p>i) true/false concerning whether authorization is needed</p> <p>True</p> <p>ii) a description of the process to obtain access to restricted content.</p> <p>False</p>	<p>Present that authorization is needed.</p> <p>Absent how to get it</p>	Shows that authorization is needed. but does not state how to get it
https://purl.org/fair-metrics/FMA2 (Metadata Longevity)	The existence of metadata even in the absence/removal of data	URL to a formal metadata longevity plan	Metadata is preserved when deleting data	Metadata is preserved when deleting data
https://purl.org/fair-metrics/FM1 (Use a Knowledge Representation Language)	Use of a formal, accessible, shared, and broadly applicable language for knowledge representation.	<p>URL to the specification of the language</p> <p>URL:</p> <p>“@prefix skos: <http://www.w3.org/2004/02/sko</p>	<p>Present</p> <p>However file needs to be created by the user.</p>	Available for metadata. not for data.

		s/core#> . “		
https://purl.org/fair-metrics/FM_1_2 (Use FAIR Vocabularies)	The metadata values and qualified relations should themselves be FAIR, for example, terms from open, community-accepted vocabularies published in an appropriate knowledge-exchange format.	UUIDs representing the vocabularies used for (meta)data @prefix locn: <http://www.w3.org/ns/locn#> .	For metadata present when RDF plugin installed for data itself needs to be created by the owner.	Available for metadata. not for data.
https://purl.org/fair-metrics/FM_1_3 (Use Qualified References)	Relationships within (meta)data, and between local and third-party data, have explicit and ‘useful’ semantic meaning	Linksets (in the formal sense) representing part or all of your resource	Possibility to make it present .	Described relationships are added. Possible relationships are not found automatically
https://purl.org/fair-metrics/FM_R1.1 (Accessible Usage License)	The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents	GUID of the license (e.g. its URL) for the data license and for the metadata license	Present name and link available.	Absent
https://purl.org/fair-metrics/FM_R1.2 (Detailed Provenance)	That there is provenance information associated with the data,	Two URLs (GUIDs). One of these URLs points to one of the	Absent Entirely dependent on the user. Not really detailed provenance.	Absent, not mandatory to add

	<p>covering at least two primary types of provenance information:</p> <ul style="list-style-type: none"> - Who/what/When produced the data (i.e. for citation) - Why/How was the data produced (i.e. to understand context and relevance of the data) 	<p>vocabularies used to describe citational provenance (e.g. dublin core). The second points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM)</p> <p>Entirely dependent on the user.</p>		
https://purl.org/fair-metrics/FM_R1.3 (Meets Community Standards)	<p>Certification, from a recognized body, of the resource meeting community standards.</p>	<p>A certification saying that the resource is compliant</p>	<p>Absent</p>	<p>Absent</p>

Overall score FAIRness CKAN:

Principle:	Score:	Missing:
Findable	4.5/5	F1B
Accessible	1/3(?)	A1.1, A1.2
Interoperable	1.5/3	
Re-usable	1/3	R1.2, R1.3

Conclusion FAIR metrics review CKAN:

High findability and possibility to become interoperable when the user puts in the effort.
Re-usability for a user is high but not for a machine.

Main concerns:

- Very dependent on the user, even in areas where the application could automate it.

FAIR metrics review Invenio

Executed by Jarno van Erp, starting January 3 2018, using
<http://cds.cern.ch/record/2298895?ln=en>

Metric identifier	What is being measured?	What must be provided?	Score Manual	Score Automated
https://purl.org/fair-metrics/FM_F1A (Identifier Uniqueness)	Whether there is a scheme to uniquely identify the digital resource.	URL to a registered identifier scheme. xmlns:dc="http://purl.org/dc/elements/1.1/"	Present. Invenio automatically creates a DOI.	Present. Invenio automatically creates a DOI.
https://purl.org/fair-metrics/FM_F1B (Identifier persistence)	Whether there is a policy that describes what the provider will do in the event an identifier scheme becomes deprecated.	A URL that resolves to a document containing the relevant policy.	Not clear if they mean that Invenio should specify this or if the identifier scheme provider should specify this. For Invenio it is absent .	Not clear if they mean that Invenio should specify this or if the identifier scheme provider should specify this. For Invenio it is absent .

<p>https://purl.org/fair-metrics/FM_F2 (Machine-readability of metadata)</p>	<p>The availability of machine-readable metadata that describes a digital resource.</p>	<p>A URL to a document that contains machine-readable metadata for the digital resource. Furthermore, the file format must be specified.</p> <p>URL:</p> <p><code>http://cds.cern.ch/record/2298895/export/xd?In=en</code></p> <p>Format specification:</p> <pre><dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:invenio="http://invenio-software.org/elements/1.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd" "></pre> <p>In the heading of metadata</p>	<p>Automatically created by Invenio. Present</p>	<p>Is automatically generated from the metadata but not for data</p>
<p>https://purl.org/fair-metrics/FM_F2</p>	<p>Whether the metadata</p>	<p>The URL of the metadata and</p>	<p>Resource is identified in</p>	<p>Resource can be identified in multiple ways</p>

<p>F3 (Resource Identifier in Metadata)</p>	<p>document contains the globally unique and persistent identifier for the digital resource.</p>	<p>the GUID of the digital resource it describes.</p> <p>URL to metadata:</p> <p>http://cds.cern.ch/record/2298895/export/xid?ln=en</p> <p>GUID:</p> <pre><dc:source>http://cds.cern.ch/record/2298895</dc:source></pre> <pre><dc:doi>10.1016/0920-5632(93)90338-7</dc:doi></pre> <pre><dc:identifier>http://cds.cern.ch/record/2298895</dc:identifier></pre> <pre><dc:identifier>oai:inspirehep.net:1644502</dc:identifier></pre>	<p>multiple ways</p> <p>Present</p>	<p>Present</p>
<p>https://purl.org/fair-metrics/FM F4 (Indexed in a searchable resource)</p>	<p>The degree to which the digital resource can be found using web-based search engines</p>	<p>The persistent identifier of the resource and one or more URLs that give search results of different search engines.</p> <p>Persistent identifier:</p> <p>10.1016/0920-5632(93)90338</p>	<p>Resource can be found through a web-based search engine.</p> <p>Present</p>	<p>Resource can be found through a web-based search engine.</p> <p>Present</p>

		<p>-7</p> <p>URL:</p> <p>https://www.google.nl/search?q=10.1016%2F0920-5632(93)90338-7&oq=10.1016%2F0920-5632(93)90338-7&aqs=chrome..69i57.281j0j1&sourceid=chrome&ie=UTF-8</p>		
<p>https://purl.org/fair-metrics/FMA1.1 (Access Protocol)</p>	<p>The nature and use limitations of the access protocol.</p>	<p>i) A URL to the description of the protocol</p> <p>Not provided but uses https</p> <p>ii) true/false as to whether the protocol is open source</p> <p>True</p> <p>iii) true/false as to whether the protocol is (royalty) free</p> <p>True</p>	<p>No URL provided to describe HTTPS. I think this is especially needed when the protocol is not common.</p>	<p>No URL provided to describe HTTPS. I think this is especially needed when the protocol is not common.</p>
<p>https://purl.org/fair-metrics/FMA1.2 (Access authorization)</p>	<p>Specification of a protocol to access restricted content.</p>	<p>i) true/false concerning whether authorization is needed</p> <p>True</p> <p>ii) a description of the process to obtain access to restricted</p>	<p>Present that authorization is needed.</p> <p>Absent how to get it</p>	<p>Shows that authorization is needed. but does not state how to get it</p>

		content. False		
https://purl.org/fair-metrics/FM_A2 (Metadata Longevity)	The existence of metadata even in the absence/removal of data	URL to a formal metadata longevity plan	Metadata is deleted when deleting data Absent.	Metadata is deleted when deleting data Absent.
https://purl.org/fair-metrics/FM_1 (Use a Knowledge Representation Language)	Use of a formal, accessible, shared, and broadly applicable language for knowledge representation.	URL to the specification of the language URL: http://www.openarchives.org/OAI/2.0/oai_dc.xsd	Present Although for data inside invenio user dependent.	Available for metadata. not for data.
https://purl.org/fair-metrics/FM_2 (Use FAIR Vocabularies)	The metadata values and qualified relations should themselves be FAIR, for example, terms from open, community-accepted vocabularies published in an appropriate knowledge-exchange format.	UUIDs representing the vocabularies used for (meta)data http://www.openarchives.org/OAI/2.0/oai_dc/	For metadata present. For data itself needs to be created by the owner.	Available for metadata. not for data.
https://purl.org/fair-metrics/FM_3 (Use Qualified References)	Relationships within (meta)data, and between local and third-party data, have explicit and 'useful' semantic meaning	Linksets (in the formal sense) representing part or all of your resource	Absent. Has optionality to some extent but not in the FAIR way	Absent. Has optionality to some extent but not in a FAIR way

https://purl.org/fair-metrics/FM_R1.1 (Accessible Usage License)	The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents	GUID of the license (e.g. its URL) for the data license and for the metadata license	Present. Although specified as rights, not as license	Present. Although specified as rights, not as license
https://purl.org/fair-metrics/FM_R1.2 (Detailed Provenance)	That there is provenance information associated with the data, covering at least two primary types of provenance information: - Who/what/When produced the data (i.e. for citation) - Why/How was the data produced (i.e. to understand context and relevance of the data)	Two URLs (GUIDs). One of these URLs points to one of the vocabularies used to describe citational provenance (e.g. dublin core). The second points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM) Creator etcetera is shown but not full provenance.	Present but it requires additional input from the user, although this input can be made mandatory	Present, meeting minimal requirements
https://purl.org/fair-metrics/FM_R1.3 (Meets Community Standards)	Certification, from a recognized body, of the resource meeting community standards.	A certification saying that the resource is compliant	Absent	Absent

Overall score FAIRness Invenio:

Principle:	Score:	Missing:
Findable	4.5/5	F1B
Accessible	0/3(?)	A1.1, A1.2, A2
Interoperable	1.5/3	I2, I3
Re-usable	0/3	R1.2, R1.2, R1.3

Conclusion FAIR metrics review Invenio:

High findability, however no license specified in machine readable format. On the CERN server there is not even a license somewhere. Zenodo seems like a better fit than pure Invenio.

Main concerns:

- No serious ways to link to other datasets
- License not present