

Appendix A

This appendix contains five subsections that detail the results of the metric analyses. Section A.1 displays the metric count for each FAIR principle relative to their declared intent. It complements what is presented and discussed in Section 4.2.1. Section A.2 reports the result of the metric classification according to the FAIR principle (or principles) we deemed the closest. Section A.3 elaborates on the A.2 results by analysing the FAIR principles co-occurrences in the metrics classified as ‘many’. Section A.4 illustrates the misalignments between the values in A.1 and A.2. Section A.2, A.3, and A.4 complement what is presented and discussed in Section 4.2.2. Finally, Section A.5 presents a taxonomy of FAIR assessment approaches by the FAIR principle, according to the declared intent of the metrics. It complements what is presented and discussed in Section 4.2.3.

A.1 Assessment metrics: declared intents

Table A.1 reports the number of metrics per tool with respect to the specific principle. The columns F, A, I, and R are used when a metric refers only to Findable, Accessible, Interoperable, and Reusable and not to a numbered/specific principle. The column ‘n/a’ is used for counting the metrics that do not declare a reference to a principle or even to a letter of the FAIR acronym. Grey rows identify tools without any reference to a specific FAIR principle.

Table A.1. FAIR assessment tools’ declared metric intent count.

tool_id	F	F1	F2	F3	F4	A	A1	A1.1	A1.2	A2	I	I1	I2	I3	R	R1	R1.1	R1.2	R1.3	n/a	total
AUT	8	0	0	0	0	4	0	0	0	0	3	0	0	0	8	0	0	0	0	0	23
CHE	0	3	4	0	0	0	0	1	1	0	0	1	2	1	0	0	1	1	2	0	17
DAT	5	0	0	0	0	5	0	0	0	0	5	0	0	0	1	0	0	0	0	0	16
ENO	0	6	4	4	3	0	0	4	4	2	0	8	4	2	0	3	2	0	1	0	47
EVA	0	4	1	1	1	0	8	2	1	1	0	4	2	6	0	1	3	2	4	0	41
EVL	0	3	2	2	4	0	1	2	2	1	0	4	3	1	0	2	2	0	2	0	31
FDB	14	0	0	0	0	14	0	0	0	0	10	0	0	0	11	0	0	0	0	0	49
FES	7	0	0	0	0	9	0	0	0	0	11	0	0	0	5	0	0	0	0	28	60
FOO	0	5	1	1	2	0	1	1	0	1	0	1	2	0	0	5	2	2	0	0	24
FRO	0	14	7	5	5	0	9	1	0	1	0	4	5	3	0	14	6	6	9	0	89
FSH	88	13	4	5	7	36	13	7	4	5	16	6	8	10	55	5	7	9	7	34	339
FUJ	0	6	4	3	3	0	8	0	0	0	0	3	3	3	0	9	3	3	9	0	57
HFI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	18
MAT	0	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	1	1	0	0	13
OFA	0	13	24	5	11	0	15	3	2	6	0	17	12	10	0	14	11	26	9	0	178
OPE	0	4	1	1	1	0	7	2	0	0	0	3	1	4	0	1	3	0	2	0	30
RDA	0	4	1	1	1	0	8	2	1	1	0	4	2	6	0	1	3	2	4	0	41
SAG	8	0	0	0	0	3	0	0	0	0	2	0	0	0	5	0	0	0	0	0	18
SAT	11	0	0	0	0	7	0	0	0	0	11	0	0	0	14	0	0	0	0	0	43
SET	14	0	0	0	0	13	0	0	0	0	10	0	0	0	9	0	0	0	0	0	46
total	155	76	54	29	39	91	71	26	16	19	68	56	44	47	108	56	44	52	49	80	1180

A.2 Assessment metrics: observed intents

Table A.2 documents the number of metrics per tool for each FAIR principle as derived from the analysis of the metrics implementation and their assignment to the FAIR principle or set of principles (column 'many'), sounding closer. The table is the basis for Section 4.2.2 and it is summarised by Figure 3.

Table A.2. FAIR assessment tools' metric count resulting from our analysis. The column many refers to metrics having more than one FAIR principle assigned, while the column none refers to metrics we consider beyond FAIR.

tool_id	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3	many	none	total
AUT	2	0	0	1	0	1	0	1	1	0	0	3	0	4	1	4	5	23
CHE	3	1	0	0	0	1	0	0	2	0	1	1	1	1	0	6	0	17
DAT	0	0	0	0	0	1	2	0	3	0	1	0	0	0	0	8	1	16
ENO	6	0	4	3	1	4	2	2	10	3	0	0	4	0	2	4	2	47
EVA	4	1	1	1	7	2	1	1	9	2	3	1	2	2	2	2	0	41
EVL	3	0	2	4	2	2	1	1	5	2	0	0	2	0	4	2	1	31
FDB	5	2	1	5	3	4	1	5	6	0	2	0	3	4	3	4	1	49
FES	4	0	0	3	2	0	3	0	6	0	1	0	2	0	0	0	39	60
FOO	5	0	1	2	0	1	0	1	1	2	0	5	2	2	0	1	1	24
FRO	10	3	4	3	6	1	0	1	9	3	3	15	6	5	5	6	9	89
FSH	31	4	6	10	14	31	5	6	26	7	9	21	28	19	9	55	58	339
FUJ	2	1	2	1	6	0	0	0	8	1	3	10	3	2	5	5	8	57
HFI	0	0	0	10	0	0	0	0	0	0	0	0	1	0	0	5	2	18
MAT	1	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0	2	13
OFA	22	24	3	11	6	3	2	1	15	5	4	18	7	23	4	5	25	178
OPE	4	1	1	1	7	2	0	0	3	1	4	1	2	0	1	2	0	30
RDA	4	1	1	1	8	2	1	1	4	2	6	1	2	1	2	4	0	41
SAG	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	2	12	18
SAT	2	2	0	0	0	0	0	1	0	0	5	1	4	7	5	15	1	43
SET	4	1	1	5	5	0	4	4	4	2	1	0	3	3	0	6	3	46
total	114	41	28	62	68	56	23	26	113	30	44	79	73	74	43	136	170	1180

A.3 Assessment metrics: FAIR principles co-occurrences

Table A.3 shows the co-occurrences among the FAIR principles we observed in the metrics of the column 'many' discussed in Section 4.2.2 (Figure 5).

Table A.3. Number of co-occurrences among FAIR principles observed in the metrics.

	F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3
F1		7		2					2			3			
F2	7								9			36		3	6

F3					2						1			
F4	2													
A1			2					13	12	2	1			12
A1.1														12
A1.2											3	4		
A2												4		
I1	2	9			13				15	2	2	7	4	35
I2					12				15		7			15
I3			1		2				2	7				2
R1	3	36			1		3		2					1
R1.1							4	4	7					
R1.2		3							4			1		2
R1.3		6			12	12			35	15	2		2	

A.4 Assessment metrics: misalignments

Table A.4 documents the number of misalignments between the originally declared intent and the observed intent. Discussion on it is in Section 4.2.2 and 5.2.2.

Table A.4. Misaligned metrics distribution: From declared (rows) to observed intent (columns). Grey shadings highlight the four FAIR areas, helping in identifying misalignments among the metrics within.

	F1	F2	F4	A1	A1.1	I1	I3	R1	R1.1	R1.2	R1.3	many	none	total
F				1	12	14		2		7	3	15	11	65
F1				5									4	9
F2				2		5						3	1	11
F3				1				2					3	6
F4						4								4
A		1						1	2	1			28	33
A1						5		4					13	22
A1.1													1	1
A1.2								1						1
A2								3		1			1	5
I	1							1					13	15
I1													4	4

	F1	F2	F4	A1	A1.1	I1	I3	R1	R1.1	R1.2	R1.3	many	none	total
I2				2		2	4					2		10
I3	10					3							3	16
R			1			1							16	18
R1							1		2	2	3		7	15
R1.1								3					3	6
R1.2								1	1			1	5	8
R1.3						5			2				9	16
n/a			10			1		2	4	2		13	48	80
total	11	1	11	11	12	40	5	20	11	13	6	34	170	345

A.5 Assessment metrics: Assessment implementations

This section presents the taxonomy of approaches followed by the metrics for assessing every FAIR principle.

F1 assessment implementations include:

- *Identifier-centred implementations*, include controls that are based on the identifier string itself and can be further divided into:
 - *syntax-based*, which leverages third-party services to verify if the identifier schema can be found in a registry (e.g. FAIRsharing) or test the identifier against a controlled list of known identifiers (e.g. PURL, DOI, W3ID).
 - *resolution-based*, verifying if the identifier can be successfully resolved (e.g. using HTTP).
 - *version-oriented*, addressing the existence of possible version-specific identifiers.
 - *metadata-specific*, verifying if metadata have their identifier, distinct from that of the digital object.
- *Metadata element-centred implementations*, verify the presence of specific metadata elements (e.g. dc:identifier, datacite:identifier).
- *Policy-centred implementations*, look at policies certifying the persistence of an identifier or documenting how changes in the identifier scheme will be managed.

F2 assessment implementations include:

- *Metadata element-centred implementations*, verify the presence of specific metadata elements (e.g. dc:creator, dc:description) and can also refer to a framework defining a minimum description for a specific object (e.g. MIRO guidelines for describing ontologies (Matentzoglou et al. 2018)).
- *Metadata value-centred implementations*, verify if a string can be found in the metadata values (e.g. if the keywords used for finding an object are all included in metadata values);
- *Format-centred*, verify the availability of metadata in specific formats (e.g. JSON, JSON-LD).

- *Identifier-centred implementations*, verify the resolution of a metadata element identifier or that the metadata element identifiers can be found in a registry (e.g. BioPortal).

F3 assessment implementations include:

- *Metadata element-centred implementations*, verify the presence of specific metadata elements (e.g. dc:source).
- *Identifier-centred implementations*, verify the resolution of the object identifier or check if metadata have their identifier – distinct from that of the digital object – and if there is an explicit link between the two.

F4 assessment implementations include:

- *Third-party service-based implementations*, verify that an object can be found by generalistic search engines (e.g. Microsoft Bing) or object-specific and domain-specific registries and repositories (e.g. RE3data).
- *Protocol-centred implementations*, verify that metadata can be harvested using a specific protocol (e.g. OAI-PMH).
- *Format-centred implementations*, verify that specific formats are used when providing metadata (e.g. JSON-LD).
- *Metadata element-centred implementations* verify that a specific metadata element is used for identifying a third-party service where the object is registered (e.g. schema:includedInDataCatalog).
- *Service-centred implementations* are linked to the FAIR assessment of registries and repositories and look at the presence of search functionality.

A1 assessment implementations include:

- *Metadata element-centred implementations* verify if specific metadata elements can be found for access rights (e.g. dc:rights) or available endpoints (e.g. sd:endpoint).
- *Metadata value-centred implementations* verify that access information is encoded using machine-readable vocabularies (e.g. COAR access rights vocabulary).
- *Identifier-centred implementations* can be divided between
 - *resolution-based*, verifying that the identifier resolves,
 - *syntax-based*, verifying that supported protocols are found in the identifier string or that content negotiation is supported (by adding different formats to the original object URI).
- *Protocol-centred implementations* verify that a specific protocol is used for accessing a digital object (e.g. HTTP) or its metadata (e.g. OAI-PMH).
- *Third-party service-based implementations* verify that a digital object can be found using a specific gateway (e.g. spark.science).

*A1.1 assessment implementations*¹ include:

- *Protocol-centred implementations* verify that a specific protocol is supported (e.g. HTTP).
- *Identifier-centred implementations* try to resolve the given identifiers or recognise their syntax and compare it against a controlled list of accepted values (e.g. ARK).

¹ A1.1 assessment approaches are all based on the use of the HTTP protocol, which is inherently free and open.

- *Metadata element-centred implementations* look for specific metadata elements describing an access point (e.g. sd:endpoint).
- *Documentation-centred implementations* require an explicit reference (e.g. a URL) to a description of the protocol used.

A1.2 *assessment implementations* include:

- *Metadata element-centred implementations* look for specific access rights metadata (e.g. odrl:hasPolicy).
- *Identifier-centred implementations* verify that identifiers resolve and check the reply.
- *Documentation-centred implementations* require an explicit reference (e.g. a URL) to a description of the process for obtaining access.

A2 *assessment implementations* include:

- *Metadata value-centred implementations* verify that a specific value defining a persistence policy is found.
- *Metadata element-centred implementations* verify that a specific metadata label is found (e.g. omv:status for ontologies).
- *Documentation-centred implementations* require that a URL to a metadata retention and preservation policy is provided.
- *Third-party service-based implementations* verify if the digital object is registered in a repository or registry (e.g. Linked Open Vocabularies for ontologies), if there is a data availability policy in the metadata provided by the service (e.g. RE3data), or if the repository provides metadata even if, given the versioning metadata, no version of the digital object results available.

I1 *assessment implementations* include:

- *Format-centred implementations* verify that metadata are provided in an accepted format (e.g. txt), structured format (e.g. JSON), or in an RDF serialisation (e.g. n3).
- *Metadata element-centred implementations* verify that a specific metadata element is found (e.g. omv:hasOntologyLanguage for ontologies).
- *Identifier-centred implementations* verify if namespaces related to RDF can be found by directly searching their strings for 'rdf' or by searching metadata value identifiers for domains linked to controlled vocabularies (e.g. vocab.getty.edu).

I2 *assessment implementations* include:

- *Metadata element-centred implementations* verify that specific metadata elements are used for finding declarations about other vocabularies (e.g. owl:imports).
- *Identifier-centred implementations* verify if specific controlled vocabularies are used by searching for domains linked to them (e.g. id.loc.gov). *Identifier-centred implementations* can be divided between.
 - *resolution-based*, verifying that the used metadata identifiers resolve.
 - *controlled list-based*, verifying the used metadata identifiers against a list of accepted semantic artefacts.
 - *third-party service-based*, verifying that the used metadata identifiers can be found in registries or repositories (e.g. BioPortal).

I3 *assessment implementations* include:

- *Identifier-centred implementations* verify if an identifier follows a specific syntax (e.g. an ORCID) or if the identifier used for referring related resources is in a controlled list of accepted identifiers.
- *Metadata element-centred implementations* verify that specific metadata elements are used to declare references to other data or metadata (e.g. dc:relation).

R1 assessment implementations include:

- *Format-centred implementations* verify if a specific format is used for serialising metadata (e.g. SHACL).
- *Metadata element-centred implementations* verify if specific metadata elements are declared (e.g. dc:type).
- *Third-party service-based implementations* verify if a metadata value can be found in a registry (e.g. if a licence can be found in the SPDX list).
- *Identifier-centred implementations* verify if a specific namespace pattern can be found in the metadata element identifiers (e.g. <http://purl.obolibrary.org/obo/HP>).
- *Metadata value-centred implementations* verify if specific information can be found in the metadata values (e.g. NIH project or award).

R1.1 assessment implementations include:

- *Metadata element-centred implementations* verify if specific metadata elements are used for declaring a licence (e.g. dc:license).
- *Identifier-centred implementations* verify if a specific licence identifier can be found or if the licence identifier can be resolved successfully.
- *Third-party service-based implementations* verify if a licence can be found in a licence registry (i.e. [SPDX](#)) or in the metadata provided by a registry (i.e. [re3data](#)).

R1.2 assessment implementations include:

- *Metadata element-centred implementations* verify if specific metadata elements are used for declaring provenance information (e.g. schema:datePublished).
- *Identifier-centred implementations* verify if a specific URI string linked to a provenance semantic artefact can be found among the declared namespaces (e.g. [W3C PROV](#)).
- *Metadata value-centred implementations* verify if specific information about a digital object is available (e.g. contact information of the creators).
- *Third-party service-based implementations* verify that specific information, such as author and title, are provided by a registry (i.e. [re3data](#)).

R1.3 assessment implementations include:

- *Third-party service-based implementations* verify if the metadata schema used can be found in a registry (e.g. BioPortal) or if it can be found among the information given by a registry (i.e. [re3data](#)). Moreover, a metric suggests using a community validator service, yet no concrete validator is advocated.
- *Format-centred implementations* verify that data or metadata are provided in a specific format (e.g. application/pdf).
- *Identifier-centred implementations* verify if the namespaces of the used metadata schemas can be found in a controlled list or if the identifier of the used metadata schema resolves successfully.
- *Metadata element-centred implementations* verify that specific metadata elements are used (e.g. omv:endorsedBy).

- *Metadata schema-centred implementations* verify that specific metadata schemas are used (e.g. DataCite metadata schema).