

EOSC-Nordic



"FAIRification of Nordic+Baltic data repositories"

Webinar April 22, 2020



EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652





Metastudy results

3

СF

Summary of metastudy findings

- Found 61 repositories with Nordic involvement from re3data.org sample
- While many of the repositories have partnered with other countries (or EU), only three (5%) has a second Nordic country among its partners. This is surprising as we expect there to be strong synergies in partnering with other Nordic countries
- Approx. 60% of the repositories do not issue PIDs, while 27% use DOI (this is the most common PID technology used)
- Almost all repositories provide unrestricted access to their metadata

BIC

- A majority (70%) of the repositories do not provide unrestricted access to all their data. Typically, some of the data is shared, while some remains restricted. For sensitive data this can be expected to some extent, but it seems to apply to repositories in all scientific disciplines
- The majority (56%) of the repositories do not employ any metadata standard
- About 80% of the repositories are not certified archives or do not follow established archive/repository standards

The state of **Open Science** in the Nordic countries





EOSC-Nordic FAIR Maturity evaluation of data repositories

Andreas O Jaunsen (NeIC / WP4 lead)



EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

WP4 members











October 5-9, 2020

Uppsala, Sweden

http://bit.ly/FAIRds-Nordic-SE

Nordic FAIR data stewardship course

- FAIRds-Nordic Norway 36 participants
- FAIRds-Nordic Denmark 31 participants
- FAIRds-Nordic Sweden 39 participants
- FAIRds-Nordic Finland ? participants







EVALUATE

EOSC-Nordic WP4: FAIR data





What is FAIR?



Problems/concerns respondents have with sharing datasets

Digital Science Report: The State of Open Data 2019



Digital Science Report: The State of Open Data 2019



What FAIR is not...



- FAIR is **not** a standard
- FAIR is not equal to 'Open' or 'Free'
 - Data are often Open but not FAIR
 - Data could be closed yet perfectly FAIR
- FAIR is **not** equal to RDF, Linked Data, or Semantic Web
- FAIR is **not** assuming that only humans can find and re-use data
- FAIR is **not** for humans only but for machines as well
- Data that are **not** FAIR are pretty 'Re-useless'.....

Source: GO-FAIR



Data as increasingly FAIR Digital Objects



Mons et al. 2017



FAIR Digital Objects



Bonino 2019



FAIR PRINCIPLES

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

Interoperable:

11. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles;

I3. (meta)data include qualified references to other (meta)data;

Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2. the protocol allows for an authentication and authorization procedure, where necessary;

A2. metadata are accessible, even when the data are no longer available;

Reusable:

R1. (meta)data are richly described with a plurality of accurate and relevant attributes;

R1.1. (meta)data are released with a clear and accessible data usage license;

R1.2. (meta)data are associated with detailed provenance;

R1.3. (meta)data meet domain-relevant community standards;

https://www.nature.com/articles/sdata201618







FAIR DATA PRINCIPLES - METADATA

Findable:

F1. metadata are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. metadata are registered or indexed in a searchable resource;

Interoperable:

I1. metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. metadata use vocabularies that follow FAIR principles;
I3. metadata include qualified references to other metadata;

Accessible:

A1. metadata are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2. the protocol allows for an authentication and authorization procedure, where necessary;

A2. metadata are accessible, even when the data are no longer available;

Reusable:

R1. metadata are richly described with a plurality of accurate and relevant attributes;

R1.1. metadata are released with a clear and accessible data usage license;

R1.2. metadata are associated with detailed provenance;

R1.3. metadata meet domain-relevant community standards;

https://www.nature.com/articles/sdata201618







FAIR DATA PRINCIPLES – DATA/DIGITAL RESOURCES

Findable:

F1. data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. data are registered or indexed in a searchable resource;

Interoperable:

I1. data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. data use vocabularies that follow FAIR principles;

13. data include qualified references to other (meta)data;

Accessible:

A1. data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2. the protocol allows for an authentication and authorization procedure, where necessary;

A2. metadata are accessible, even when the data are no longer available;

Reusable:

R1. data are richly described with a plurality of accurate and relevant attributes;

R1.1. data are released with a clear and accessible data usage license;

R1.2. data are associated with detailed provenance;

R1.3. data meet domain-relevant community standards;

https://www.nature.com/articles/sdata201618







FAIR PRINCIPLES - TECHONOLOGY-RELATED

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

Interoperable:

In (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles;

I3. (meta)data include qualified references to other (meta)data;

Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2. the protocol allows for an authentication and authorization procedure, where necessary;

A2. metadata are accessible, even when the data are no longer available;

Reusable:

R1. (meta)data are richly described with a plurality of accurate and relevant attributes;

R1.1. (meta)data are released with a clear and accessible data usage license;

R1.2. (meta)data are associated with detailed provenance;

R1.3. (meta)data meet domain-relevant community standards;







FAIR Maturity evaluation

Why evaluate repositories?

- Data are and will be distributed in small and typically domain specific data repositories (not in large data silos)
- Wish to help such repositories identify possible areas of improvements of their service to become FAIRer
- Raise awareness of FAIR practices and the importance of using <u>machine-actionable</u> metadata
- Contribute to FAIR uptake across region and thereby the premise for better reuse of the data

Measuring the FAIR Maturity of repositories

- We consider a MANUAL approach to be both time-consuming, prone to biases and not (very) reproducible
- The preferred method is to perform AUTOMATED evaluations using a well defined set of test criteria / metrics (FAIR Maturity indicators)
- <u>Wilkinson et al. 2018</u> (doi:10.1038/sdata.2018.118) provides a framework and metrics for measuring FAIRness of data and Mark Wilkinson's gen2 tests (22 tests) and evaluator tool: <u>https://fairsharing.github.io/FAIR-Evaluator-FrontEnd</u> provides the best **current** tool to achieve this
- The FAIR Maturity evaluator provides <u>efficiency</u>, <u>scalability</u> and <u>reproducibility</u>



FAIR Maturity indicators measure aspects of the FAIR principles

FAIR Maturity indicators



	Metric name	Principle association	Principle description
1	UNIQUE IDENTIFIER	F1	(Meta)data are assigned a globally unique and persistent identifier
2	IDENTIFIER PERSISTENCE	F1	(Meta)data are assigned a globally unique and persistent identifier
3	DATA IDENTIFIER PERSISTENCE	F1	(Meta)data are assigned a globally unique and persistent identifier
4	STRUCTURED METADATA	F2	Data are described with rich metadata (defined by R1 below)
5	GROUNDED METADATA	F2	Data are described with rich metadata (defined by R1 below)
6	DATA IDENTIFIER EXPLICITLY IN METADATA	F3	Metadata clearly and explicitly include the identifier of the data they describe
7	METADATA IDENTIFIER EXPLICITLY IN METADATA	F3	Metadata clearly and explicitly include the identifier of the data they describe
8	SEARCHABLE IN MAJOR SEARCH ENGINE	F4	(Meta)data are registered or indexed in a searchable resource
9	USES OPEN FREE PROTOCOL FOR DATA RETRIEVAL	A1.1	The protocol is open, free, and universally implementable
10	USES OPEN FREE PROTOCOL FOR METADATA RETRIEVAL	A1.1	The protocol is open, free, and universally implementable
11	DATA AUTHENTICATION AND AUTHORIZATION	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
12	METADATA AUTHENTICATION AND AUTHORIZATION	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
13	METADATA PERSISTENCE	A2	Metadata are accessible, even when the data are no longer available
14	METADATA KNOWLEDGE REPRESENTATION LANGUAGE (WEAK)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
15	METADATA KNOWLEDGE REPRESENTATION LANGUAGE (STRONG)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
16	DATA KNOWLEDGE REPRESENTATION LANGUAGE (WEAK)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
17	DATA KNOWLEDGE REPRESENTATION LANGUAGE (STRONG)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
18	METADATA USES FAIR VOCABULARIES (WEAK)	12	(Meta)data use vocabularies that follow FAIR principles
19	METADATA USES FAIR VOCABULARIES (STRONG)	12	(Meta)data use vocabularies that follow FAIR principles
20	METADATA CONTAINS QUALIFIED OUTWARD REFERENCES	13	(Meta)data include qualified references to other (meta)data
21	METADATA INCLUDES LICENSE (STRONG)	R1.1	(Meta)data are released with a clear and accessible data usage license
22	METADATA INCLUDES LICENSE (WEAK)	R1.1	(Meta)data are released with a clear and accessible data usage license
		R1.2	(Meta)data are associated with detailed provenance
		R1.3	(Meta)data meet domain-relevant community standards

https://doi.org/10.1038/s41597-019-0184-5

https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/





Evaluate resources

Evaluate resources FAIRness against Collections of Maturity Indicator Tests



This application is driven by the FAIRmetrics and the FAIRsharing groups. We recognize the support of the DBCLS BioHackathon series during which much of the back-end code was prototyped. Licensed under MIT.



This application is driven by the FAIRmetrics and the FAIRsharing groups. We recognize the support of the DBCLS BioHackathon series during which much of the back-end code was prototyped. Licensed under MIT.









(cc)) BY



Evaluation methodology

Repository selection

- Repository must have Nordic+Baltic relation (contain data from region)
- Sample is not exhaustive, but hopefully representative
- Exclude repositories containing only publications/articles
- Select repositories that are considered relevant sources of data for research related re-use
- Repository must identify datasets by globally unique identifiers (GUID) in order to be selected for evaluation



Dataset selection

- If repository satisfies the above selection criteria we proceed to perform DO/dataset selection
- Randomly (and manually) select N=10 datasets from each repository, scattering the selection across time submitted and across scientific domains
- Exclusively use URIs as dataset/DO identifier (may change this)
- We take any dataset to be representative of the repository in which it resides



Dataset selection

- If repository satisfies the above selection criteria we proceed to perform DO/dataset selection
- Randomly (and manually) select N=10 datasets from each repository, scattering the selection across time submitted and across scientific domains
- Exclusively use URIs as dataset/DO identifier (may change this)

 We take any dataset to be representative of the repository in which it resides



Does a single dataset evaluation reliably indicate repository FAIR maturity level?



Consistency test

- 1. Evaluate a few repositories from the sample, one from each score category (low, medium, high)
- Perform the FAIR maturity evaluations for each repository using N=10



Consistency test

- 1. Evaluate a few repositories from the sample, one from each score category (low, medium, high)
- Perform the FAIR maturity evaluations for each repository using N=10

_			_				DS									
F	Α		R	FAIR	Avg	Var	1	2	3	4	5	6	1	8	9	10
50%	40%	57%	0%	37%	9.0	0.00	<u>9</u>									
25%	40%	0%	0%	16%	4.0	0.00	<u>4</u>									
63%	40%	71%	100%	68%	13.4	0.52	<u>14</u>	<u>13</u>	<u>13</u>	<u>13</u>	<u>14</u>	<u>13</u>	<u>13</u>	<u>14</u>	<u>14</u>	<u>13</u>
75%	80%	71%	100%	82%	15.3	4.00	<u>17</u>	<u>17</u>	<u>16</u>	<u>4</u>	<u>17</u>	<u>16</u>	<u>17</u>	<u>16</u>	<u>17</u>	<u>16</u>



Aggregated results



100 repositories, 72 evaluated



FAIR

100 repositories, 72 evaluated







Histogram of R-SCORE



Dataset result (example)

DO evaluations

repolD \Xi	Evaluation result string =	F-score \Xi	A-score \Xi	I-score 😇	R-score =	FAIR score =	Succeded tests / Total tests =
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001110011110110011100	50.00%	80.00%	71.43%	0.00%	59.09%	(13:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001110011110110011100	50.00%	80.00%	71.43%	0.00%	59.09%	(13:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
27	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
54	1001100001010110011100	37.50%	40.00%	71.43%	0.00%	45.45%	(10:22)
26	1001110011110110011111	50.00%	80.00%	71.43%	100.00%	68.18%	(15:22)
26	1001110011110110111111	50.00%	80.00%	85.71%	100.00%	72.73%	(16:22)
26	1001110011110111111111	50.00%	80.00%	100.00%	100.00%	77.27%	(17:22)
26	1001110011110110011111	50.00%	80.00%	71.43%	100.00%	68.18%	(15:22)
26	1001110011110110111111	50.00%	80.00%	85.71%	100.00%	72.73%	(16:22)
26	1001110011110110011111	50.00%	80.00%	71.43%	100.00%	68.18%	(15:22)
26	1001110011110110011111	50.00%	80.00%	71.43%	100.00%	68.18%	(15:22)
26	1001110011110111111111	50.00%	80.00%	100.00%	100.00%	77.27%	(17:22)
26	1001110011110110011111	50.00%	80.00%	71.43%	100.00%	68.18%	(15:22)
26	1001110011110111111111	50.00%	80.00%	100.00%	100.00%	77.27%	(17:22)
24	100000001010000000000	12.50%	40.00%	0.00%	0.00%	13.64%	(3:22)
24	100000001010000000000	12.50%	40.00%	0.00%	0.00%	13.64%	(3:22)
24	100000001010000000000	12.50%	40.00%	0.00%	0.00%	13.64%	(3:22)
24	100000001010000000000	12.50%	40.00%	0.00%	0.00%	13.64%	(3:22)

Fri, 17 Apr 2020 16:26:00 + Test of: https://plos.figshare	-0000 e.com/articles/_Test_res	sults_of_group_differences	_in_cognitive_performance	GUID: https://plos.figshare.com/articles/_Test_results_of_group_d ifferences_in_cognitive_performance_domains_between_low e_d_vs_high_SCC_groups_8224_/1080323 Date: Fri, 17 Apr 2020 16:26:00 +0000
				FAIR Metrics Gen2 - Data Knowledge Representation Language (strong)FAIR Metrics Gen2 - Data Knowledge Representation Language (weak)FAIR Metrics Gen2 - Metadata contains qualified outward references)
F Metrics	A Metrics	I Metrics	R Metrics	FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong) FAIR Metrics Gen2 - Metadata Knowledge Representation Language (weak) FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (strong) FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)

Fri, 17 Apr 2020 16:26:00 +0000

Test of: https://plos.figshare.com/articles/_Test_results_of_group_differences_in_cognitive_performance_dHTTP_Accept header {"Accept"=>"*/*"}.



Using the output from this URL for the next few tests... INFO: Found type of content when resolving https://plos.figshare.com/ndownloader/files/1561141 using WARN: parser could not be found. INFO: Metadata may be embedded, now searching using the Apache 'tika' tool. INFO: The message body is being examined by Apache Tika INFO: The response from Apache Tika is being parsed INFO: entering Tika parser - sample of input <x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe . INFO: Tika executed successfully (this doesn't necessarily mean that it discovered any metadata...) INFO: Metadata may be embedded, now searching using the 'Distiller' tool. INFO: Cached data is already parsed. Returning INFO: Metadata may be embedded, now searching using the extruct' tool. INFO: Using 'extruct' to try to extract metadata from return value (message body) of https://s3-eu-west-1.amazonaws.com/pstorage-plos-3567654/1561141/Table 3.xls. WARN: extruct threw an error Failed to extract rdfa, raises 'utf-8' codec can't decode byte 0xd0 in position 0: invalid continuation byte when attempting to parse return value (message body) of https://s3-eu-west-1.amazonaws.com/pstorage-plos-3567654/1561141/Table 3.xls. INFO: The GUID of the data appears to be a URL. SUCCESS: The data was found to have some Linked Data content. FAIR Metrics Gen2 - Data Knowledge Representation

FAIR Metrics Gen2 - Metadata contains qualified outward references)

Language (weak)

FAIR Metrics Gen2 - Metadata Knowledge Representation



Some details

- 714 datasets evaluated for this study
- 103.7 hours execution time for the full sample
- NOTE: indicator test "Metadata Identifier Explicitly in Metadata" only accepts EXACT match
- NOTE: indicator test "Searchable in Major Search Engine" was disabled for this run due to lack of valid license for Bing

FAIR Maturity indicators



	Metric name	Principle association	Principle description
1	UNIQUE IDENTIFIER	F1	(Meta)data are assigned a globally unique and persistent identifier
2	IDENTIFIER PERSISTENCE	F1	(Meta)data are assigned a globally unique and persistent identifier
3	DATA IDENTIFIER PERSISTENCE	F1	(Meta)data are assigned a globally unique and persistent identifier
4	STRUCTURED METADATA	F2	Data are described with rich metadata (defined by R1 below)
5	GROUNDED METADATA	F2	Data are described with rich metadata (defined by R1 below)
6	DATA IDENTIFIER EXPLICITLY IN METADATA	F3	Metadata clearly and explicitly include the identifier of the data they describe
7	METADATA IDENTIFIER EXPLICITLY IN METADATA	F3	Metadata clearly and explicitly include the identifier of the data they describe
8	SEARCHABLE IN MAJOR SEARCH ENGINE	F4	(Meta)data are registered or indexed in a searchable resource
9	USES OPEN FREE PROTOCOL FOR DATA RETRIEVAL	A1.1	The protocol is open, free, and universally implementable
10	USES OPEN FREE PROTOCOL FOR METADATA RETRIEVAL	A1.1	The protocol is open, free, and universally implementable
11	DATA AUTHENTICATION AND AUTHORIZATION	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
12	METADATA AUTHENTICATION AND AUTHORIZATION	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
13	METADATA PERSISTENCE	A2	Metadata are accessible, even when the data are no longer available
14	METADATA KNOWLEDGE REPRESENTATION LANGUAGE (WEAK)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
15	METADATA KNOWLEDGE REPRESENTATION LANGUAGE (STRONG)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
16	DATA KNOWLEDGE REPRESENTATION LANGUAGE (WEAK)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
17	DATA KNOWLEDGE REPRESENTATION LANGUAGE (STRONG)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
18	METADATA USES FAIR VOCABULARIES (WEAK)	12	(Meta)data use vocabularies that follow FAIR principles
19	METADATA USES FAIR VOCABULARIES (STRONG)	12	(Meta)data use vocabularies that follow FAIR principles
20	METADATA CONTAINS QUALIFIED OUTWARD REFERENCES	13	(Meta)data include qualified references to other (meta)data
21	METADATA INCLUDES LICENSE (STRONG)	R1.1	(Meta)data are released with a clear and accessible data usage license
22	METADATA INCLUDES LICENSE (WEAK)	R1.1	(Meta)data are released with a clear and accessible data usage license
		R1.2	(Meta)data are associated with detailed provenance
		R1.3	(Meta)data meet domain-relevant community standards

https://doi.org/10.1038/s41597-019-0184-5



Histogram of FAIR Maturity test passes



Early/prelim results from 48 tested URIs and 12 matching DOIs



Mirror, mirror, on the wall... who's the FAIRest of them all?

repolD	D)ata-se	Platform	F-score	A-score	I-score	R-score	FAIR	Sigma	Sigma (F)	Sigma (A)	Sigma (I)	Sigma (R)	CTS	DSA	WDS 3	LARIN
	2 1	0	Dspace	37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000	X			X
	3 1	0		30.00%	40.00%	62.86%	0.00%	40.00%	0.164	0.065	0.000	0.100	0.000				
	4	5		12.50%	40.00%	0.00%	0.00%	13.04%	0.000	0.000	0.000	0.000	0.000				
	6 1	0	META-SHAPE	12 50%	40.00%	0.00%	0.00%	13 64%	0.129	0.000	0.000	0.009	0.000	v			×
	7 1	0	HETA SHARE	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	^			^
	8 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	9 8		Dataverse	50.00%	80.00%	71.43%	0.00%	59.09%	0.000	0.000	0.000	0.000	0.000				
	10 1	1	NESSTAR	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	x			
	11 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	13 1	0	Dspace	20.00%	40.00%	8.57%	0.00%	19.09%	0.259	0.121	0.000	0.138	0.000	х			Х
	16 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	18 1	1	Dataverse	50.00%	80.00%	77.14%	0.00%	60.91%	0.138	0.000	0.000	0.138	0.000	Х			
	19 1	0	Nesstar	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	20 1	0	Dataverse	50.00%	80.00%	71.43%	0.00%	59.09%	0.117	0.000	0.000	0.117	0.000				
	24 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000		X		X
	25 1	0	P1 1	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	26	0	Figshare	50.00%	80.00%	82.80%	100.00%	/1.82%	0.131	0.000	0.000	0.131	0.000	×			
	28 1	0		40.00%	40.00%	29 57%	0.00%	21 9 24	0.221	0.000	0.109	0.000	0.000	^			
	29 1	a		12 50%	40.00%	0 0.0%	0.00%	13 64%	0.000	0.000	0.000	0.000	0.000				
	30 1	a		12.50%	40.00%	0.00%	0 00%	13 64%	0.000	0.000	0.000	0.000	0.000				
	32 1	1	IPT	37.50%	40.00%	68,83%	100.00%	53.72%	0.058	0.000	0.000	0.058	0.000			х	
	35 1	0	21113D	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	39 1	0		23.75%	40.00%	21.43%	30.00%	27.27%	1.009	0.181	0.000	0.345	0.483				
	41 1	0		25.00%	40.00%	14.29%	0.00%	22.73%	0.282	0.132	0.000	0.151	0.000				
	42 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	45 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	47 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	49 4			18.75%	40.00%	7.14%	25.00%	20.45%	0,768	0.125	0.000	0.143	0.500				
	52 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	54 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	55 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	57 1	0		17.50%	40.00%	11.43%	0.00%	19.09%	0.346	0.105	0.000	0.241	0.000	_			
	60 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	_			
	62 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	_			
	63 1	0		35.00%	40.00%	25.71%	0.00%	30.00%	0.169	0.079	0.000	0.090	0.000				
	64	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000	~			v .
	65	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000	X			x
	00 1	0	Fischara	12.50%	40.00%	77 1 49	100.00%	13.04%	0.000	0.000	0.000	0.000	0.000				
	60 1	0	rigshare	12 50%	10 00%	0.00%	0.00%	12 648	0.100	0.000	0.000	0.100	0.000				
	71 1	0		15 0.0%	40.00%	2 86%	10 00%	16.36%	0.000	0.000	0.000	0.000	0.000				
	72 1	0		37.50%	40.00%	65 31%	100.00%	52.60%	0.976	0.000	0.000	0.076	0.000				
	73 7			37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	76 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	79 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	82 1	3	CKAN	50.00%	80.00%	79.76%	0.00%	61.74%	0.129	0.000	0.000	0.129	0.000				
	84 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	85 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	87 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	94 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
2	100 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	103 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
-	106 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
9	108 4		CKAN	50.00%	80.00%	100.00%	50.00%	72.73%	0.577	0.000	0.000	0.000	0.577				
23	113 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
10	114 1	0		37.50%	40.00%	/1.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
1	115	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	120	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
22	120 1	9		20.00%	40.00%	25 71*	0.00%	30.00%	0.259	0.121	0.000	0.138	0.000				
1	125 1	0		30 00%	40.00%	50 00%	0.00%	35 919	0.109	B 121	0.000	B 345	0.000	-			
3	127 1	0		37 50%	40.00%	28.57%	0.00%	31.82%	8,000	0.000	0 000	0.000	0.000				
2	129 1	0		37 50%	40 00%	28 57%	0.00%	31 82%	0 000	0 000	0 000	8 888	0 000				
	130	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
14	131 1	0		12.50%	40,00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
-	132 1	0	Dataverse	50,00%	80,00%	69.84%	0.00%	58.59%	0.048	0.000	0.000	0.048	0.000				
2	133 1	0		37.50%	40.00%	42.86%	0.00%	36.36%	0.000	0.000	0.000	0.000	0.000				
e	134 1	0		37.50%	40.00%	57.14%	0.00%	40.91%	0.000	0.000	0.000	0.000	0.000				
	135 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
5	136 1	0	figshare	46.25%	68.00%	64.29%	70.00%	59.09%	8, 938	0.060	0.193	0.193	0.483				

epolD	D	ata-se	Platform	F-score	A-score	I-score	R-score	FAIR	Sigma	Sigma (F)	Sigma (A)	Sigma (I)	Sigma (R)	CTS	DSA	WDS 20	ARN
	2 1	0	Dspace	37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000	х		х	(
	3 1	0		30.00%	40.00%	62.86%	0.00%	40.00%	0.164	0.065	0.000	0.100	0.000				
	4 1	5		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	5 1	0		21.25%	40.00%	10.00%	0.00%	20.00%	0.129	0.060	0.000	0.069	0.000				-
	6 1	0	META-SHARE	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	X	_	X	8
	0 1	0		12.50%	40.00%	0.00%	0.00%	13.04%	0.000	0.000	0.000	0.000	0.000				
	0 0	U	Dataverse	50 00%	80.00%	71 42%	0.00%	50 0.0%	0.000	0.000	0.000	0.000	0.000				
	10 1	1	NESSTAR	12 50%	40.00%	0 00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	x		_	
	11 1	0	incoorran.	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	~			
	13 1	0	Dspace	20.00%	40.00%	8.57%	0.00%	19.09%	0.259	0.121	0.000	0.138	0.000	x		x	(
	16 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	18 1	1	Dataverse	50.00%	80.00%	77.14%	0.00%	60.91%	0.138	0.000	0.000	0.138	0.000	х			
	19 1	0	Nesstar	12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	20 1	0	Dataverse	50.00%	80.00%	71.43%	0.00%	59.09%	0.117	0.000	0.000	0.117	0.000				
	24 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000		X	Х	1
	25	-		10 500	10.000	0.000	0.000	1.0.0.00	2.000	0.000	2.000		0.000			-	
	26 1	0	Figshare	50.00%	80.00%	82.86%	100.00%	71.82%	0.131	0.000	0.000	0.131	0.000			_	
	21 1	0		40.00%	40.00%	/1.43%	0.00%	40.10/4	0.221	0.003	0.103	0.000	0.000	^			
	28 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	29 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000			-	
	30	1	TPT	27 50%	40.00%	68 939	100.00%	52 728	0.000	0.000	0.000	0.000	0.000			x	
	35 1	0	*11	12 50%	40.00%	0 00%	0.00%	13 64%	0.058	0.000	0.000	0.058	0.000			^	
	39 1	0		23.75%	40.00%	21.43%	30.00%	27.27%	1 640	0.181	0.000	0.345	0.483				
	41 1	0		25,00%	40.00%	14.29%	0.00%	22.73%	0.282	0.132	0.000	0.151	0.000				
	42 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	45 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	47 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	49 4			18.75%	40.00%	7.14%	25.00%	20.45%	0.768	0.125	0.000	0.143	0.500				
	52 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	54 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	55 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	57 1	0		17.50%	40.00%	11.43%	0.00%	19.09%	0.346	0.105	0.000	0.241	0.000				
	60 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	62 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	63 1	0		35.00%	40.00%	25.71%	0.00%	30.00%	0.169	0.079	0.000	0.090	0.000				
	64 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	65 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000	X		X	(
	0.0								0 000	0.000	0 000		0 000		-		
	68 1	0	Figshare	50.00%	80.00%	77.14%	100.00%	70.00%	0.100	0.000	0.000	0.100	0.000				_
	09 1	0		12.00%	40.00%	0.00%	0.00%	13.04%	0.000	0.000	0.000	0.000	0.000				-
	71 1	0		15.00%	40.00%	2.80%	10.00%	10.30%	0.480	0.079	0.000	0.090	0.316	-			
	72 7	0		37.50%	40.00%	00.31%	0.00%	21 92%	0.010	0.000	0.000	0.070	0.000				
	76 1	a		37.50%	40.00%	28.57%	0.00%	31.02%	6 000	0.000	0.000	0.000	0.000				
	79 1	0		12 50%	40.00%	0.00%	0.00%	13.64%	0.000	0 000	0.000	0.000	0.000				
	82 1	3	CKAN	50.00%	80.00%	79.76%	0.00%	61.74%	0,129	0.000	0.000	0,129	0.000				
	84 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0,000	0.000	0.000	0.000	0.000				
	85 1	0		37.50%	40,00%	71,43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	87 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	94 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	100 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	103 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	100						2.22		2.000			2.000	2.000				
	108 4		CKAN	50.00%	80.00%	100.00%	50.00%	72.73%	0,577	0.000	0.000	0.000	0.577				
	113 1	0		12.50%	40.00%	0.00%	0.00%	13.04%	0.000	0.000	0.000	0.000	0.000				
	114 1	0		37.50%	40.00%	71.43%	0.00%	45.45%	0.000	0.000	0.000	0.000	0.000				
	115 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	116 1	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	120 1	0		20.00%	40.00%	8.57%	0.00%	19.09%	0.259	0.121	0.000	0.138	0.000				
	122 1	0		35.00%	40.00%	25.71%	0.00%	30.00%	0.169	0.079	0.000	0.090	0.000	-			
	125 1	0		30.00%	40.00%	50.00%	0.00%	35.91%	0.466	0.121	0.000	0.345	0.000			-	
	12/ 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000			_	
	129 1	0		37.50%	40.00%	28.57%	0.00%	31.82%	0.000	0.000	0.000	0.000	0.000				
	130	0		12.50%	40.00%	0.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000				
	131	0	Datawaras	12.50%	40.00%	60.00%	0.00%	13.64%	0.000	0.000	0.000	0.000	0.000	-		-	
	132 1	0	vataverse	37.50	40.00%	43.84%	0.00%	36.59%	0.048	0.000	0.000	0.048	0.000			-	
	124	0		37.50%	40.00%	42.80%	0.00%	40.01%	0.000	0.000	0.000	0.000	0.000				
	135 1	0		12 50%	40.00%	0.00%	0.00%	13 649	0.000	0.000	0.000	0.000	0.000				
	136 1	0	figshare	46.25%	68 00%	64 29%	70 00%	59 099	8 536	0.000	0 192	0 193	8 482				



Software







Software platforms...

	ΤοοΙ	Implementation	Cost	Platform	Installation	User interface	ΑΡΙ
ArchivesSpace	ArchivesSpace	Download	Free	Lin Mac Win	Moderate	Web	Yes
🚩 ckan	CKAN	Download Web Service	Free Subscription	Lin	Complex	Web	Yes
CONTENTdm	CONTENTdm	Download Service	Subscription	Lin Win	Simple	Web	Yes
DataFlow	DataBank	Download	Free	Lin	Complex	CL Web	Yes
D SPACE	DSpace	Download	Free	Lin Mac Win	Moderate	Web	Yes
eprints repeatory software	EPrints	Download	Free	Lin Mac Win	Moderate	Web	Yes
Fedora Commons	Fedora	Download	Free	Lin Mac Win	Complex	CL Web	Yes





Conclusions



Highlights

- Collected 136 regional digital repositories from eight countries and evaluated 100 of them.
- Evaluation based on machine-actionable metadata, provided DO has GUID (identifier).
- Evaluation of a *small* number of datasets (N=10) within a repository is typically sufficient to determine a repository FAIR score. However, larger samples (N=100) should be considered. Listing of ALL datasets in a repository should be a *generic* feature!
- Evaluations consist of harvesting metadata/data from GUID by resolving all links within the DO langing page. This takes 5-20 minutes per dataset. Parallelised evaluations for speedup using 10 workers and automatic execution and results extraction from Google sheets using Google scripts.
- Evaluation of multiple datasets (N=10) to estimate an <u>average</u> FAIR Maturity score for the repository (code published as open source)
- Streamlined FAIR Maturity evaluation of datasets is a scalable approach to determine FAIRness implementation

Recommendations Eosc

- All datasets should be identified by a globally unique identifier (GUID), preferably a persistent identifier (PID)
- Repositories should register on <u>re3data.org</u> to increase discoverability
- Employ the concept of FAIR digital object for published datasets (cf. "Metdata Identifier Explicitly in Metadata" and "Data Identifier Explicitly in Metadata")
- Make use of linked
- State under what license agreement the dataset is provided, using one of the standard "license" predicates/keys

FAIR Maturity indicators



	Metric name	Principle association	Principle description
1	UNIQUE IDENTIFIER	F1	(Meta)data are assigned a globally unique and persistent identifier
2	IDENTIFIER PERSISTENCE	F1	(Meta)data are assigned a globally unique and persistent identifier
3	DATA IDENTIFIER PERSISTENCE	F1	(Meta)data are assigned a globally unique and persistent identifier
4	STRUCTURED METADATA	F2	Data are described with rich metadata (defined by R1 below)
5	GROUNDED METADATA	F2	Data are described with rich metadata (defined by R1 below)
6	DATA IDENTIFIER EXPLICITLY IN METADATA	F3	Metadata clearly and explicitly include the identifier of the data they describe
7	METADATA IDENTIFIER EXPLICITLY IN METADATA	F3	Metadata clearly and explicitly include the identifier of the data they describe
8	SEARCHABLE IN MAJOR SEARCH ENGINE	F4	(Meta)data are registered or indexed in a searchable resource
9	USES OPEN FREE PROTOCOL FOR DATA RETRIEVAL	A1.1	The protocol is open, free, and universally implementable
10	USES OPEN FREE PROTOCOL FOR METADATA RETRIEVAL	A1.1	The protocol is open, free, and universally implementable
11	DATA AUTHENTICATION AND AUTHORIZATION	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
12	METADATA AUTHENTICATION AND AUTHORIZATION	A1.2	The protocol allows for an authentication and authorisation procedure, where necessary
13	METADATA PERSISTENCE	A2	Metadata are accessible, even when the data are no longer available
14	METADATA KNOWLEDGE REPRESENTATION LANGUAGE (WEAK)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
15	METADATA KNOWLEDGE REPRESENTATION LANGUAGE (STRONG)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
16	DATA KNOWLEDGE REPRESENTATION LANGUAGE (WEAK)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
17	DATA KNOWLEDGE REPRESENTATION LANGUAGE (STRONG)	11	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
18	METADATA USES FAIR VOCABULARIES (WEAK)	12	(Meta)data use vocabularies that follow FAIR principles
19	METADATA USES FAIR VOCABULARIES (STRONG)	12	(Meta)data use vocabularies that follow FAIR principles
20	METADATA CONTAINS QUALIFIED OUTWARD REFERENCES	13	(Meta)data include qualified references to other (meta)data
21	METADATA INCLUDES LICENSE (STRONG)	R1.1	(Meta)data are released with a clear and accessible data usage license
22	METADATA INCLUDES LICENSE (WEAK)	R1.1	(Meta)data are released with a clear and accessible data usage license
		R1.2	(Meta)data are associated with detailed provenance
		R1.3	(Meta)data meet domain-relevant community standards

https://doi.org/10.1038/s41597-019-0184-5

Thank you