

EOSC Nordic

FAIR uptake in the Nordics and Baltics

RDA Deutschland Tagung 25.02.2022

Mari Kleemola, FSD/Tampere University





In this presentation

- EOSC Nordic project shortly
- Measuring FAIR scores of metadata in Nordic and Baltic repositories
- Preliminary results
- Lessons learned



EOSC Nordic key facts

EOSC-Nordic aims to facilitate the coordination of European Open Science Cloud (EOSC) relevant initiatives within the Nordic and Baltic countries. The project aims to exploit synergies to achieve greater harmonisation at policy and service provisioning across these countries, in compliance with EOSC agreed standards and practices.

- September 2019 August 2022
- 24 project participants from 10 countries
- Project Coordinator: <u>NeIC</u> / <u>NordForsk</u>
- One of five regional INFRAEOSC 5b projects



https://www.eosc-nordic.eu/

EOSC NORDIC



OBJECTIVE 2

Increase the discoverability of Nordic & Baltic services. Extend and expand their use by making them accessible through the EOSC portal

OBJECTIVE 3

Promote and support the uptake of FAIR data practices FAIR data and certification schemas across the Nordics

OBJECTIVE 4

Accelerate the progress and attractiveness of EOSC by piloting & delivering innovative solutions developed and tested in a useful and functional cross-border environment

Support coordination, harmonisation and alignment of Nordic and Baltic national policies and practices related to the provision of horizontal research data services with EOSC

OBJECTIVE 1



Provide a Knowledge Hub to deliver training and technical support to new service providers and communities willing to engage with EOSC during and after the project lifetime MAIN OBJECTIVES

4







EOSC-Nordic WP4 goals

FAIR Digital Objects



IDENTIFIERS Persistent and unit Digitor Objects should such as a DOT or difference to other indirector biological

projects (RAUCH) 1.4

STANDARDS do Open, documenter Digeo Objects more formats. This exceeds formats are easier to formats are easier to the code use to prove

METADATA Contextual docum In order for Digital Co.

Trustworthy Digital Repositories



Digitalbevaring.dk

IN



WP4 FAIR metrics activities

Excluded repositories that don't assign a GUID to each individual dataset



Surveyed the Nordics and Baltics for 100(+) research repositories

Evaluation of 10 datasets (metadata records) per repository Monthly automated evaluations, in the beginning with the FAIR Evaluation Service (Wilkinson's tool), now with the F-UJI tool (FAIRsFAIR)

The workflow

=	FAIR evaluation production (F-UJI) ☆ ⊡ ⊘ File Edit View Insert Format Data Tools Extensions Help Last edit was made 8 days ago by Eosc Nordic														hare (
IC.		p 👎 100% 🕶 \$ % .0 .	.00 <u>123</u> -	Arial 👻	10 - B	<u>।</u> २ <u>A</u>	♦. ⊞ ∃	E • E • .	<u>t</u> + ÷ + 17	- GD 🕀	ili 🍸 👻	Σ -									
10	×	<i>f</i> x 63																			
	A	В	С	D	E	F	G	н	1	J	ĸ	L	М	N	0	Р	Q	R	S	т	
	repoID	datasetID	GUID does not resolve 404 (0) 200 (773)	Evaluation result s	F-score (7)	A-score (3)	I-score (4)	R-score (10)	FAIR score	Succeded tests / Total tests	Status Error (5)	Analyze start	t Analyze end time	Total time for analyzing	Ŋ	Descriptiv e core metadata elements F2 (376)	Contains data identifer F3 (98)	Metadata can be retrieved programm atically F4 (321)	Access level and conditions A1 (103)	Knowledg e Represent ation Language I1-01 (196)	R
2	27	https://snd.gu.se/en/catalogue/stu dy/snd0020	200	1110100111101110	64.29%	33.33%	100.00%	30.00%	52.08%	(12.5:24)	Ready	27-Jan-2022,	27-Jan-2022, 07:10:25	0:00:18	FALSE	1	0	1	0	1	
3	27	https://snd.gu.se/en/catalogue/stu dy/snd1115	200	1110100111101110	64.29%	33.33%	100.00%	30.00%	52.08%	(12.5:24)	Ready	27-Jan-2022,	27-Jan-2022, 07:10:54	0:00:16		1	0	1	0	1	
ł.	27	https://snd.gu.se/en/catalogue/stu dy/snd1080	200	1110100111101110	50.00%	33.33%	100.00%	30.00%	47.92%	(11.5:24)	Ready	27-Jan-2022,	27-Jan-2022, 07:11:27	0:00:20		1	0	1	0	1	

- Analysis is getting started in GoogleSheets (2 modes, with and without DataCite metadata)
- GoogleScripts run in the background
- One analysis takes around 20 seconds, for ca. 800 datasets it takes 4-5 hours
- Summary for the entire sample is generated automatically
- More data-analysis needs manual work



FAIR scores of repositories*



98 repositories (75 evaluated)

* August 2021, incl. DataCite metadata



FAIR score evolution, preliminary results



- DataCite metadata gives added FAIR-value
- Especially I and R scores are affected
- General (slight) increase over time
- Affected by change of version in F-UJI

Supporting FAIR uptake

We have

- published the detailed results to repositories and communities so that they can use the feedback as a guideline to improve FAIRness
- offered recommendations
- held webinars/workshops with 500+ attendees in total:
 - Apr 2020 First assessment hackathon Initial exercise
 - Nov 2020 Step 1 Focus on PIDs
 - Feb 2021 Step 2 Split between Data and Metadata
 - Apr 2021 Step 3 Generic Metadata
 - Oct 2021 Step 4 Domain-Specific Metadata
 - Dec 2021 M4M event for climate community
 - Feb 2022 Step 5 Value and Limitations of FAIR evaluators

repoiD -	Name	Date -	Platform +	F-score T	A-score -	I-score +	R-score -	FAIR =	Sigma -	ilgma (F 👻	iigma (A 🕂	Sigma (I 🔻	ilgma (F 🛨	1	1	1	1
2	CLARIN-DK	40	Dspace	50.00%	66.67%	9.72%	22.22%	34.30%	0.036	0.072	0.000	0.124	0.000	х			х
2	DDA	28		57 14%	33 33%	50 00%	33 33%	43 48%	0 073	0 293	0 000	0 000	0 000				_
3	DDA	20		57.14%	33.33%	50.00%	33.33%	45.40%	0.075	0.233	0.000	0.000	0.000				_
4	Det Kgl. bibliotek	20		30.00%	13.33%	5.00%	7.22%	14.57%	0.118	0.387	8.294	0.131	0.154				
6	Kielipankki	10	META-SHARE	14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000	х			x
	Data Canadas Dantal Mila	10		71 478	CC . C7W		22.228	47 029	0 000	0.000	0.000	0.000	0.000				_
1	Data Service Portai Alla	10		71.43%	00.07%	25.00%	33.33%	47.03%	0.000	0.000	0.000	0.000	0.000				_
8	Fairdata IDA	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
9	NMBU dataverseNO	16	Dataverse	85.72%	66.67%	37.50%	43.05%	58.15%	0.050	0.148	0.000	0.129	0.038				
	100		NECCESIO.		10.000			16 749	0.007								-
10	NSD	20	NESSTAR	37.86%	13.33%	0.00%	8.89%	10.74%	0.08/	0.283	0.168	0.000	0.112	x			_
11	HUNT Databank	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
13	CLARING Bergen Center repr	20	Danace	30 00%	20 00%	7 50%	5 67%	15 65%	0 096	0 153	9 313	8 118	0 104	x			x
15	center bergen center rep		ospace		20.00%			10.000	0.050	0.155	0.010	0.110	0.104				^
16	Språkbanken	10		14.29%	33.33%	0.00%	22.22%	17.39%	0.000	0.000	0.000	0.000	0.000				_
17	ESS Data	9		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
19	TROLLing	22	Datavorse	75 00%	56 674	27 50%	25 554	48 91%	0 103	0.240	0.244	0 160	0 160	×			
10	INGLING	22	Dataverse	75.00%	50.07%	27.30%	55.55%	40.51%	0.105	0.240	0.244	0.100	0.100	^			-
19	EED	10	Nesstar	14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
20	UiT Open Research Data Dat	20	Dataverse	85.72%	66.67%	32.50%	44.44%	57.83%	0.047	0.147	0.000	0.118	0.000				
	Production and the second			17.00				E 259	0.010		0.000		0.000				
24	sprakbanken	13		17.59%	0.00%	0.00%	0.00%	5.35%	0.010	0.003	0.000	0.000	0.000		×		^
25	Lund University Humanities	20		21.43%	0.00%	0.00%	0.00%	6.52%	0.018	0.073	0.000	0.000	0.000				
26	su figshare.com	20	Figshare	51.43%	45.00%	10.00%	29.44%	34.78%	0.098	0.287	0.196	0.126	0.136				
20	Source and the second		1 agonare		101001			10 000	0.050								-
27	SND	20		64.29%	50.00%	42.50%	38.89%	48.69%	0.077	0.220	0.171	0.118	0.057	х			_
28	ICES data portals	8		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
20	TASPAR	10		14 20%	22 224	0 007	11 11%	13 04%	0 000	0 000	0 000	0 000	0 000				
29	JASPAR	10		14.23%	33.33%	0.00%	11.11%	13.04%	0.000	0.000	0.000	0.000	0.000				_
30	STRING	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
32	GBIF	22	IPT	59.74%	66.67%	25.00%	38.89%	46.44%	0.047	0.180	0.000	0.000	0.057			x	
	UD A	10		14 000	0.000	0.000	0.000	4 259	0.000	0.000	0.000	0.000	0.000			-	
39	nra.	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				_
41	Fishbase	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
45	ISIG	10		14,29%	0.001	8.892	0.001	4.35%	9,999	0,000	0.000	8,888	0,000				
								4.000	0.000								-
47	GERDA	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
49	ACTRIS	8		46.43%	33.34%	12.50%	16.67%	27.17%	0.139	0.312	8.356	0.231	0.178				
52	NPDC	20		37 86%	35 00%	13 75%	17 22%	25 22%	0 127	A 259	9 366	0 151	0 178				
52	NFDC	20		37.00%	35.00%	13.75%	17.22%	23.22%	0.127	0.233	0.300	0.151	0.170				-
54	Bolin Centre Database	12		61.90%	33.33%	25.00%	22.22%	36.23%	0.028	0.111	0.000	0.000	0.000				
55	SMHI open data	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
	NTRD teching	20		44.20%	20.00*	0.00*	11 110	20 429	0.002	0.011	0.100	0.000	0.114				
57	NIRD AFCHIVE	20		44.29%	20.00%	0.00%	11.11%	20.43%	0.095	0.311	0.108	0.000	0.114				_
60	GTN-P Database	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
62	UNITE	28		35.72%	33.34%	6.25%	16.67%	22.83%	0.114	0.220	8.342	8,111	0.171				
02								17 200									-
63	Estonian Biocentre Public	10		14.29%	33.33%	0.00%	22.22%	17.39%	0.000	0.000	0.000	0.000	0.000				_
64	DataDOI	18		61.90%	62.97%	2.78%	27.78%	38.40%	0.068	0.208	0.108	0.081	0.116				
65	CELR META-SHARE	28		35 72%	33 33%	0 002	5 56%	17 39%	0 057	0 220	0 000	0 000	0 057	x			¥
05	CEER TIETA STORE	20		55.724			5.50%		0.057	0.220			0.057	~			^
66	AHEAD	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
68	USN RDA	10	Figshare	28.57%	33.33%	0.00%	22.22%	21.74%	0.000	0.000	0.000	0.000	0.000				
74	LOAR	10		61 654	62 164	2 629	26 214	37 76%	0 077	0.242	0 152	0 070	0.095				
71	LUAR	19		01.05%	03.10%	2.03%	20.31%	37.70%	0.0//	0.243	0.155	0.079	0.005				_
72	AIDA Data Hub	20		65.71%	43.33%	23.75%	30.00%	41.52%	0.080	0.188	0.219	0.099	0.103				
73	OoG Institute's data	10		31,43%	10.00%	0.00%	6.67%	13.48%	0.084	0.276	0.161	0.000	0.107				
	The				22.245			10 70%	0.000								
76	747	20		35.72%	33.34%	1.25%	11.11%	19.70%	0.093	0.0/3	8.342	0.055	0.114				_
78	B2SHARE	12	Invenio	38.10%	27.78%	0.00%	13.89%	20.65%	0.118	0.275	8.343	0.000	0.172				
79	DH	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
																	-
80	NLL	10		25.71%	53.34%	40.00%	35.55%	35.65%	0.101	0.060	0.281	0.211	0.187				_
	RTU RIS																
84		10		14.29%	0.00%	25.00%	11.11%	13.04%	0.000	0.000	0.000	0.000	0.000				
85	FinBIF	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
07	SARV	10		14 208	33 334	0.00*	11 118	13 049	0 000	0.000	0.000	0 000	0.000				
6/	unit	10		14.29%	33.33%	0.00%	11.11%	13.04%	0.000	0.000	0.000	0.000	0.000				_
92	SSRI	15		42.86%	26.67%	0.00%	13.33%	21.74%	0.131	0.362	0.338	0.000	0.169				
94	IINH	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
400	OcarDP	20		25 705	20 000	11.054	15	22 028	0.110	0.000		0.000	0.100				
100	usarud	20		35.72%	30.00%	11.25%	15.55%	22.03%	0.110	0.228	0.323	0.128	0.163				_
104	Bird	20		30.00%	40.00%	0.00%	6.67%	16.96%	0.074	0.222	0.137	0.000	0.137				
106	Migration Institute of Fir	10		14,29%	0.001	0.00%	0.001	4.35%	0,000	0,000	0,000	8,868	0,000				
								AE OF	0.000								-
108	Musiikkiarkisto	4	CKAN	28.57%	66.67%	50.00%	50.00%	45.65%	0.016	0.000	0.000	0.000	0.064				
109	SLS	14		14.29%	23.81%	0.00%	7.94%	10.56%	0.041	0.000	0.156	0.000	0.052				
140	SweFree	10		14 208	0.000	0.004	0.00%	4 35%	0 000	0.000	0.000	0 000	8 999				
113	Sweet Led	10		14.29%	0.00%	0.004	0.004	4.55%	0.000	0.000	0.000	0.000	0.000				-
114	Metabolic Atlas	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
115	SEAD	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
440	NOW	10		14 000	0.000	0.000	0.000	4 259	0.000	0.000	0.000	0.000	0 000				
116	NUM	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
117	SNM Digital Assets	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
120	GEUS	10		14.29%	0.007	0.007	0.007	4.35%	0.000	0.000	0.000	8.889	0.000				
120				11.234	0.004	0.004	0.004	1.30%	0.000	0.000	0.000	0.000	0.000				-
123	LARM	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
127	Garamantas	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
100	MMB	10		14 208	0 000	0 004	0.00%	4 35%	0 000	0.000	0.000	0 000	0 000				
129	nno	10		14.29%	0.00%	0.00%	0.00%	4.33%	0.000	0.000	0.000	0.000	0.000				
130	PlutoF	20		35.72%	31.67%	3.75%	16.11%	21.96%	0.111	0.220	0.333	0.092	0.167				
131	MIDAS	10		14.29%	0.00%	0.00%	0.00%	4.35%	0.000	0.000	0.000	0.000	0.000				
	1990	-	2			20.000	35	EA OF	0.000		0.000						-
132	NMUC	20	vataverse	85.72%	66.67%	32.50%	35.55%	54.35%	0.048	0.147	0.000	0.118	0.046				
133	IINH BIOTA	10		14.29%	33.33%	0.00%	22.22%	17.39%	0.000	0.000	0.000	0.000	0.000				
134	TCOS	20		21,43%	0.001	0.001	8.88%	6.52%	0,018	0.073	0.000	8,000	0,000				
1.34	*			21.438	0.004	0.004	0.004	DO FO	0.010		0.000	0.000	0.000				-
135	CESSDA DC	14		34.70%	19.05%	14.29%	12.70%	20.50%	0.139	0.335	0.313	0.234	0.208				
136	DTU data	10	figshare	28.57%	33.33%	0.00%	22.22%	21.74%	0.000	0.000	0.000	0.000	0.000				
107	CLAPTN TS	20	CLARTN	50 005	63.248	21.058	22 228	35 97%	0.020	0.073	0.102	0.002	0.000				
137	CERTER 13	20	CLARIN	30.00%	33.34%	21.20%	22.22%	00.01A	0.039	0.073	0.103	0.032	0.000				
130																	

F-UJI evaluator (lates

https://docs.google.c om/spreadsheets/d/ 1MBTMXb5SIeaBKi yEzImgJtQtxNS1zZT qTvTk9A2vqdc

🖊 NORDIC

Lessons learned



- FAIR is still poorly supported in Nordic/Baltic repositories: 1/3 of the datasets do not pass any test, while 2/3 score less than 33%
 - lack of machine-actionable metadata
 - lack of GUID
- Software and/or metric changes influence the results
- URL deprecation is a problem
- Focussing on metadata is crucial; usually metadata is openly available but data not
- Repositories should do FAIRification because it benefits their users, not because of FAIR tests
- Do not worry if you do not reach 100% FAIRness, it is more important to understand the results and limitations to improve practices
- Communities are important
- We need to think about keeping (meta)data FAIR in the long term, too.



Sources and further reading

- Webinar: Value and limitations of FAIR assessment tools 8.2.2022
- Mihai, Hannah & Tuomas J. Alaterä (2022). <u>Real world experience</u> with evaluating repositories
- Jaunsen, Andreas (2021). <u>EOSC-Nordic WP4 FAIR assessments &</u> preliminary results
- Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, Adil Hasan, Josefine Nordling, & Pauli Assinen. (2020).
 D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). Zenodo. <u>https://doi.org/10.5281/zenodo.4045402</u>
- EOSC Association Task Force on FAIR Metrics and Data Quality



EOSC

Thank you for your attention!