

2. Webinar on user experience with FAIR evaluation tools and services

11 October 2022, 2. Webinar on user experience with FAIR evaluation tools and services

 cessda.eu

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Licence: CC-BY 4.0

Using F-UJI to assess domain-specific FAIRness of research data - implementing SSH-specific metrics and testing

Robert Huber

rhuber@uni-bremen.de

Project background: FAIR Data Assessment Pilots



FAIR assessment implementation comprises the development of two main components – **assessment metrics** and **tool**.

Priority Recommendations

- Rec. 8: Facilitate automated processing
- Rec. 12: Develop metrics for FAIR Digital Objects

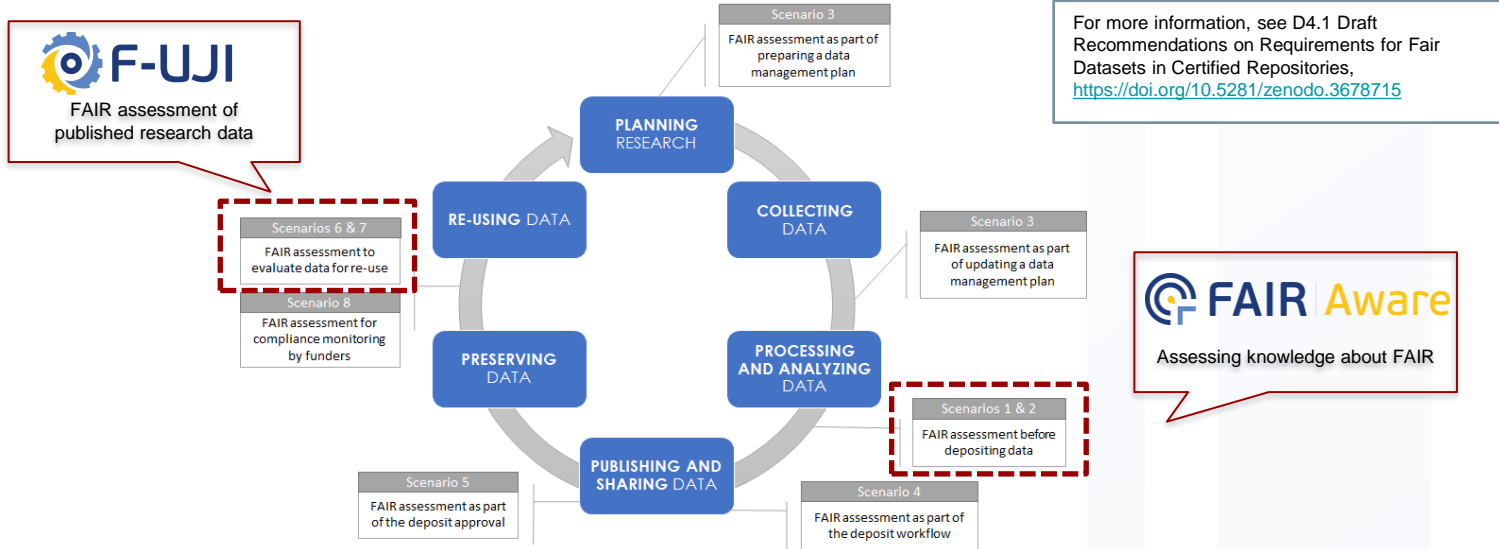
Supporting Recommendations

- Rec. 25: Implement FAIR metrics to monitor uptake



European Commission Expert Group on FAIR Data. 2018. 'Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data.' <https://doi.org/10.2777/1524>

Assessment Scenarios



Research data lifecycle; figure adapted from (Mosconi et al., 2019) and scenarios of FAIR assessment of datasets therein.

Object Assessment Metrics



FAIRSFair
Fostering Fair Data Practices in Europe

[About](#) [FAIR Support](#) [FAIR Landscape](#) [Tools & Software](#) [Events](#) [Project Outputs](#) [Outreach](#)

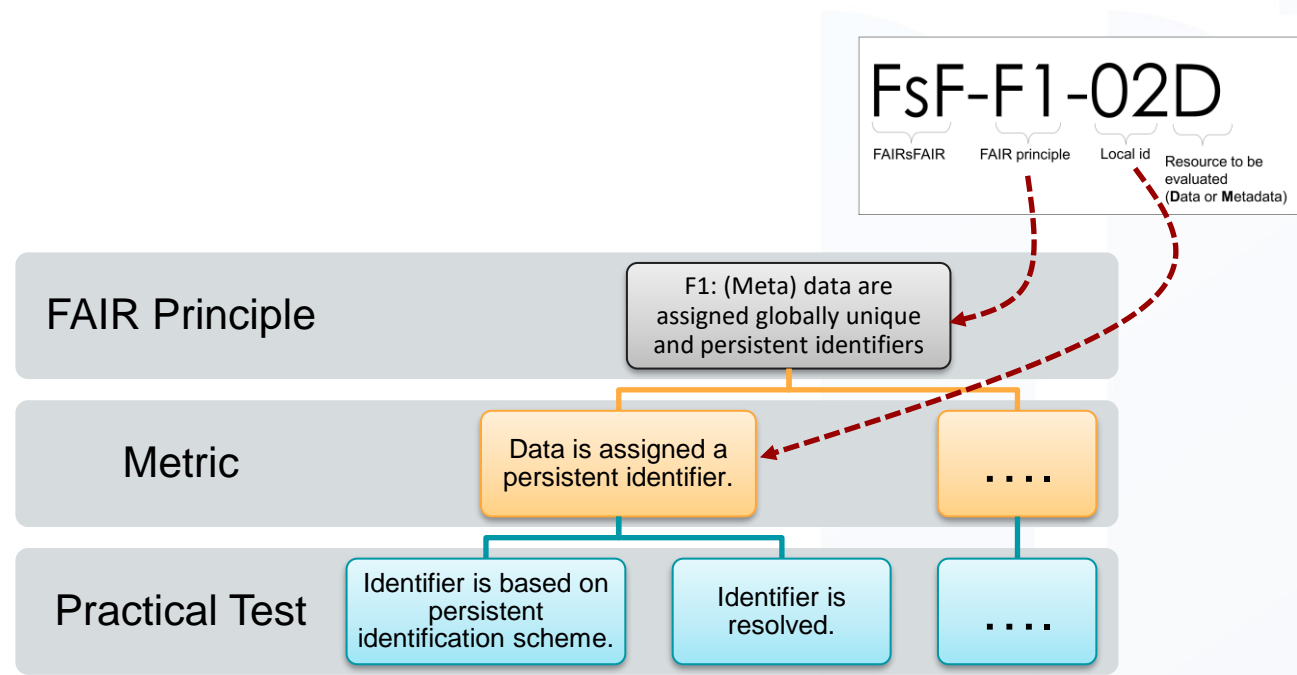
While FAIR principles may apply to any digital objects, we are concerned with the subset of digital objects: research data that are collected, measured, or created for purposes of scientific analysis.

- ✓ **FsF-F1-01D - Data is assigned a globally unique identifier**
- ✓ **FsF-F1-02D - Data is assigned a persistent identifier**
- ✓ **FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability**
- ✓ **FsF-F3-01M - Metadata includes the identifier of the data it describes**
- ✓ **FsF-F4-01M - Metadata is offered in such a way that it can be retrieved by machines**
- ✓ **FsF-A1-01M - Metadata contains access level and access conditions of the data**
- ✓ **FsF-A2-01M - Metadata remains available, even if the data is no longer available**
- ✓ **FsF-I1-01M - Metadata is represented using a formal knowledge representation language**
- ✓ **FsF-I1-02M - Metadata uses semantic resources**
- ✓ **FsF-I3-01M - Metadata includes links between the data and its related entities**
- ✓ **FsF-R1-01MD - Metadata specifies the content of the data**
- ✓ **FsF-R1.1-01M - Metadata includes license information under which data can be reused**
- ✓ **FsF-R1.2-01M - Metadata includes provenance information about data creation or generation**
- ✓ **FsF-R1.3-01M - Metadata follows a standard recommended by the target research community of the data**
- ✓ **FsF-R1.3-02D - Data is available in a file format recommended by the target research community**

Please login & comment below citing in the subject line the Metric Identifier No. you are referring to - e.g. "FsF-R1.3-01M"

v0.5

From Principles to Practical Tests



Huber, Robert, Cepinkas, Linas, Davidson, Joy, Herterich, Patricia, L'Hours, Hervé, Mokrane, Mustapha, von Stein, Ilona, & Verburg, Maaïke. (2021). D4.5 Report on FAIR Data Assessment Toolset and Badging Scheme (V1.0_DRAFT). Zenodo. <https://doi.org/10.5281/zenodo.5336159>



<https://github.com/pangaea-data-publisher/fuji>

<https://www.f-uji.net>

F-UJI – An Automated FAIR Data Assessment Tool

F-UJI 1.0.0 OAS3

[/fuji/api/v1/openapi.json](#)

A Service for Evaluating Research Data Objects Based on [FAIRsFAIR Metrics](#).

This work was supported by the [FAIRsFAIR](#) project (H2020-INFRAEOSC-2018-2020 Grant Agreement 831558).

Contact the developer
MIT License
[Find out more about F-UJI](#)

Servers
/fuji/api/v1 Authorize

FAIR object FAIRness assessment of a data object

POST /evaluate

FAIR metric FAIRsFAIR assessment metrics

GET /metrics Return all metrics and their definitions

Code **Details**

200

Response body

```

{
  "metric_identifier": "Fsf-F1-020",
  "metric_name": "Persistent identifier",
  "output": {
    "pid": "https://doi.org/10.1594/PANGAEA.902845",
    "pid_scheme": "doi",
    "resolvable_status": true,
    "resolved_url": "https://doi.pangaea.de/10.1594/PANGAEA.902845"
  },
  "passed": true,
  "score": {
    "earned": 1,
    "total": 1
  },
  "test_debug": [
    "INFO: Persistence identifier scheme - doi",
    "INFO: Retrieving page https://doi.org/10.1594/PANGAEA.902845",
    "INFO: Request status code - 200",
    "INFO: Found HTML page"
  ]
}

```

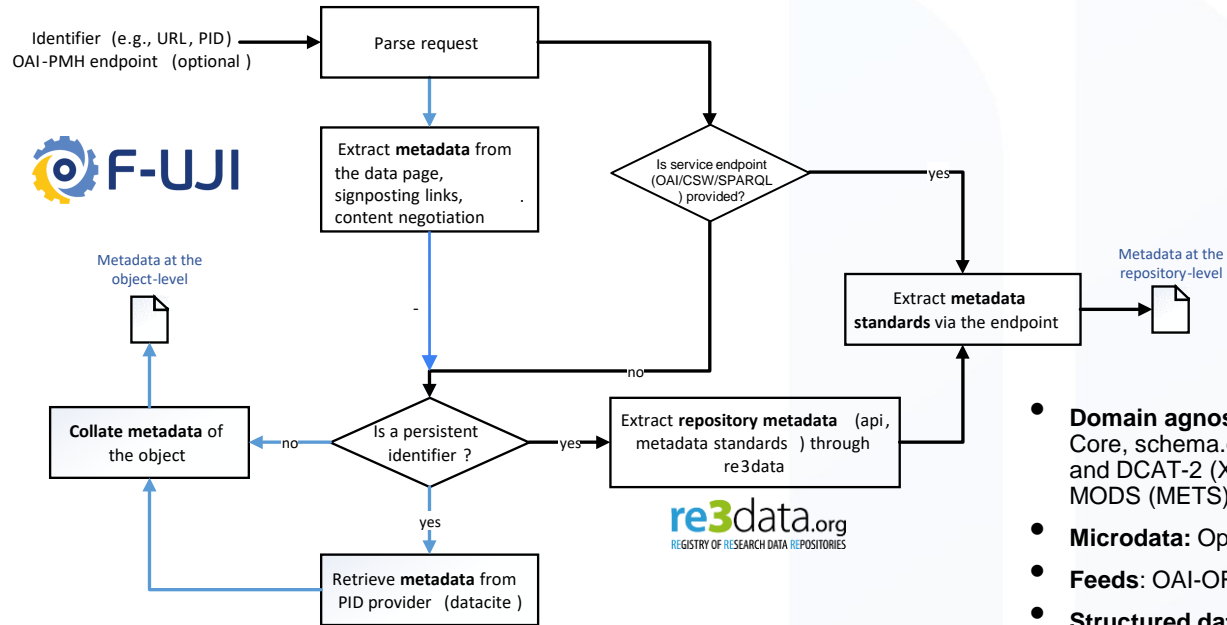
Response headers

```

content-length: 5116
content-type: application/json
date: Fri, 24 Apr 2020 17:14:06 GMT
server: Werkzeug/1.0.0 Python/3.7.6

```


High level data flow



- **Domain agnostic standards:** Dublin Core, schema.org/Dataset, DataCite, and DCAT-2 (XML, RDF, or JSON), MODS (METS) (XML)
- **Microdata:** OpenGraph, RDFa
- **Feeds:** OAI-ORE, atom or GeoRSS
- **Structured data:** RDF, RDFa, JSON-LD, turtle etc
- **Domain specific:** DDI Codebook, ISO 19115 (ISO 19139) EML



F-UJI – An Automated FAIR Data Assessment Tool

Assessment Results:

Evaluated Resource:

Data for: Bar chart of ceramic building material quantities by context type and Bar chart of ceramic building material MSW by context type and Ceramic building materials by context type (excluding Phase 6).

FAIR level: initial

Resource PID/URL: <https://doi.org/10.17863/CAM.14473>

DataCite support: enabled

Metric Version: metrics_v0.4

Metric Specifications: <https://doi.org/10.5281/zenodo.4081213>

Software version: v1.3.8

Download assessment results: [JSON](#)

Save and share assessment results:

Saved assessments:

- 2021-09-17
- 2021-09-20
- 2021-09-21
- 2021-09-21

Summary:

	Score earned:	Fair level:
Findable:	5 of 7	moderate
Accessible:	1.5 of 3	initial
Interoperable:	1 of 4	initial
Reusable:	3 of 10	initial

Report:

Findable

FsF-F1-01D - Data is assigned a globally unique identifier. ✓ ↓

FsF-F1-02D - Data is assigned a persistent identifier. ✓ ↑

FAIR level: 3 of 3 advanced

Score: 1 of 1

Output:

```
{
  "pid": "https://doi.org/10.17863/CAM.14473",
  "pid_scheme": "doi",
  "resolvable_status": true,
  "resolved_url": "https://www.repository.cam.ac.uk/handle/1810/288289"
}
```

Metric tests:	Test:	Test name:	Score:	Maturity:	Result:
	FsF-F1-02D-1	Identifier follows a defined persistent identifier syntax	0.5	1	✓
	FsF-F1-02D-2	Persistent identifier is resolvable	0.5	3	✓

Debug messages:

Level:	Message:
INFO	PID schemes-based assessment supported by the assessment service - [ark:/, arxiv/, bioproject/, biosample/, doi/, ensemble/, genome/, gnd/, handle/, iaid/, pmid/, pmcid/, puri/, refseq/, sra/, uniprot/, urn/]
INFO	Retrieving page - http://doi.org/10.17863/CAM.14473 as text/html, application/xhtml+xml, application/xml;q=0.5, text/xml;q=0.5, application/pdf;q=0.5
INFO	Content negotiation accept=text/html, application/xhtml+xml, application/xml;q=0.5, text/xml;q=0.5, application/pdf;q=0.5, status=200
INFO	Found HTML page!
INFO	Object identifier active (status code = 200)
SUCCESS	Persistence identifier scheme = doi

FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability. ✓ ↓

FsF-F3-01M - Metadata includes the identifier of the data it describes. ? ↓

FsF-F4-01M - Metadata is offered in such a way that it can be retrieved programmatically. ✓ ↓

<https://www.f-uji.net>



FAIR-IMPACT

Expanding FAIR solutions across EOSC

- We will extend and adapt the FAIRsFAIR data object assessment metrics and F-UJI tool to be more **disciplinary-context aware**
- **Discipline-aware metrics** [D5.1] and tests [MS5.4] will be developed with use case partners, domain data repositories, research infrastructures and e-infrastructures.
- A reference collection of test datasets [MS5.1] will be provided for verification and benchmarking of FAIR assessment tools' results.
- Pilots [MS5.7] will test FAIR assessment tools including additional **disciplinary-extended** tests (WP2).

FAIR-Impact integral concept: Use cases


- **life sciences** (EMBL-EBI & UNIMAN)
- **photon and neutron science** (UKRI-STFC)
- **agri-food & environmental sciences** (INRAE, LifeWatch, CNR),
- **social sciences and humanities** (**CESSDA**, UESSEX-UKDS, NSD)



Figure 2: Outcomes achieved via use case co-creation

Towards SSH FAIR metrics

- Identify core disciplinary standards
 - Ontologies
 - Metadata standards
 - Metadata properties
 - Data standards (formats)
- Define discipline specific metrics
- Identify use case specific reference datasets



FAIRSFAR Fostering Fair Data Practices in Europe

About FAIR Support FAIR Landscape Tools & Software Events Project Outputs Outreach

¹While FAIR principles may apply to any digital objects, we are concerned with the subset of digital objects: research data that are collected, measured, or created for purposes of scientific analysis.

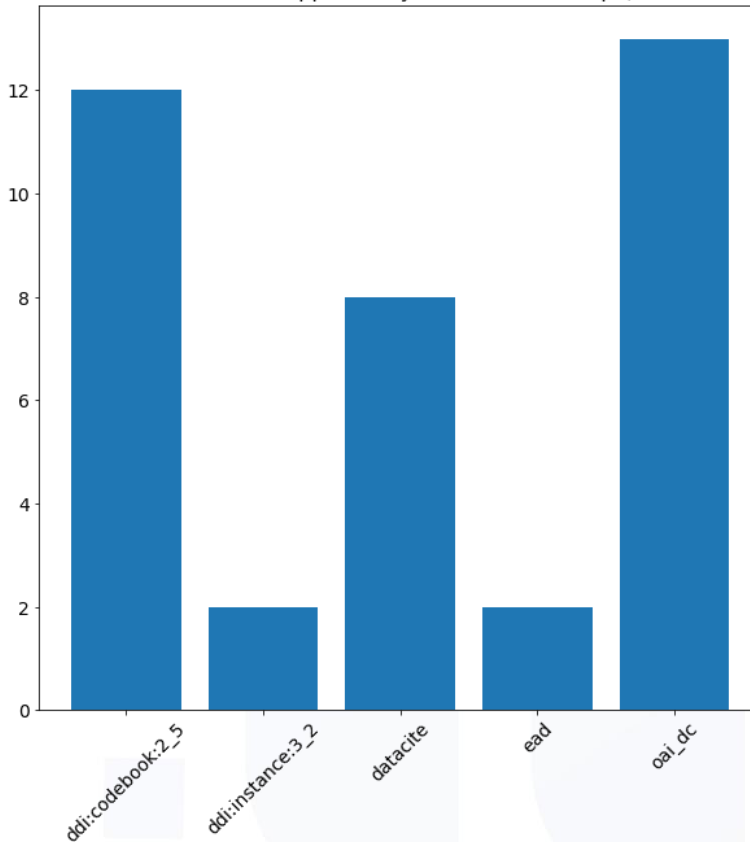
- ▼ FsF-F1-01D - Data is assigned a globally unique identifier
- ▼ FsF-F1-02D - Data is assigned a persistent identifier
- ▼ **FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability**
- ▼ FsF-F3-01M - Metadata includes the identifier of the data it describes
- ▼ FsF-F4-01M - Metadata is offered in such a way that it can be retrieved by machines
- ▼ FsF-A1-01M - Metadata contains access level and access conditions of the data
- ▼ FsF-A2-01M - Metadata remains available, even if the data is no longer available
- ▼ FsF-I1-01M - Metadata is represented using a formal knowledge representation language
- ▼ **FsF-I1-02M - Metadata uses semantic resources**
- ▼ FsF-I3-01M - Metadata includes links between the data and its related entities
- ▼ FsF-R1-01MD - Metadata specifies the content of the data
- ▼ FsF-R1.1-01M - Metadata includes license information under which data can be reused
- ▼ FsF-R1.2-01M - Metadata includes provenance information about data creation or generation
- ▼ **FsF-R1.3-01M - Metadata follows a standard recommended by the target research community of the data**
- ▼ **FsF-R1.3-02D - Data is available in a file format recommended by the target research community**

Please login & comment below citing in the subject line the Metric Identifier No. you are referring to - e.g. "FsF-R1.3-01M"

Approach

- Investigate existing implementation using interfaces (OAI-PMH)
- Collect FAIR-Implementation Profiles (FIP)
 - <https://ds-wizard.org/fair>

CESSDA metadata standards supported by more than one repo, found via OAI-PMH



SSH specific metrics definition

- SSH specific metrics
- Metric appendix –SSH
- Draft as Google doc
- Community feedback

2.1 SSH Community Metadata Standard

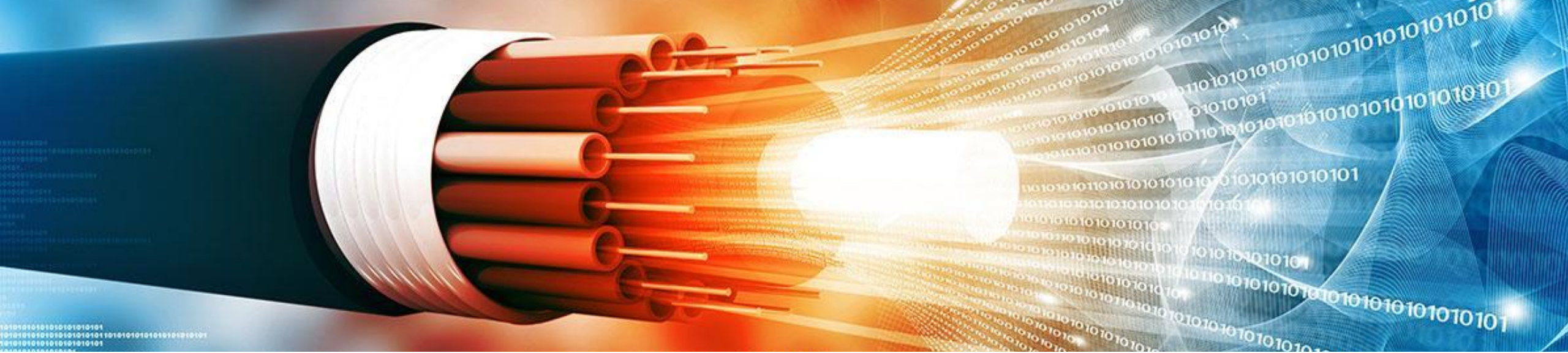
FIELD	DESCRIPTION
Metric Identifier	FsF-R1.3-01M-SSH
Metric Name	Metadata follows a standard recommended by the SSH research community of the data.
Description	<p>In addition to core metadata required to support data discovery (covered under metric FsF-F2-01M), metadata to support data reusability should be made available following community-endorsed metadata standards.</p> <p>For social sciences several well established metadata standards exist in particular the family of standards defined by the DDI (Data Documentation Initiative) Alliance.</p> <p>A SSH repository should support the following standards</p> <ul style="list-style-type: none"> ● DDI Lifecycle 3.3 ● DDI Codebook 2.5 <p>for data set level metadata description.</p>
FAIR Principle	R1.3. (Meta)data meet domain-relevant community standards
CoreTrustSeal Alignment	R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data
ASSESSMENT	
Requirement(s)	<ul style="list-style-type: none"> ● Data identifier (IRI, URL, PID) ● Standards registered and verified in re3data repository record ● Metadata access via: <ul style="list-style-type: none"> ○ Provision endpoints such as OAI-PMH offering metadata in community specific format ○ Community specific metadata links provided in the landing page via signposting or typed links or via content negotiation.

Opportunities

- FAIR Implementation Framework
- Task 2.2, 2.1
- March 2023 First Open Call
- max 10.000€
- fair-impact.eu

Thank you !

publisher/f...ssues



How FAIR evaluation tools can help increasing the FAIRness of a research data repository

CESSDA Metadata Office: Follow-up webinar on user experience with FAIR evaluation tools and services - 11.10.22

Hannah Mihai



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

FAIR ecosystem

- Research data will not become nor stay FAIR by magic.
- FAIRness requires curation and care both now and in the long term.

“FAIR Digital Objects can only exist in a FAIR ecosystem, comprising key data services that are needed to support FAIR. These include ... stewardship and repositories ...”

“Repositories offer databases and data services and should be certified to ensure trust.”

European Commission expert group on FAIR data (2018). Turning FAIR into reality. <https://doi.org/10.2777/1524>



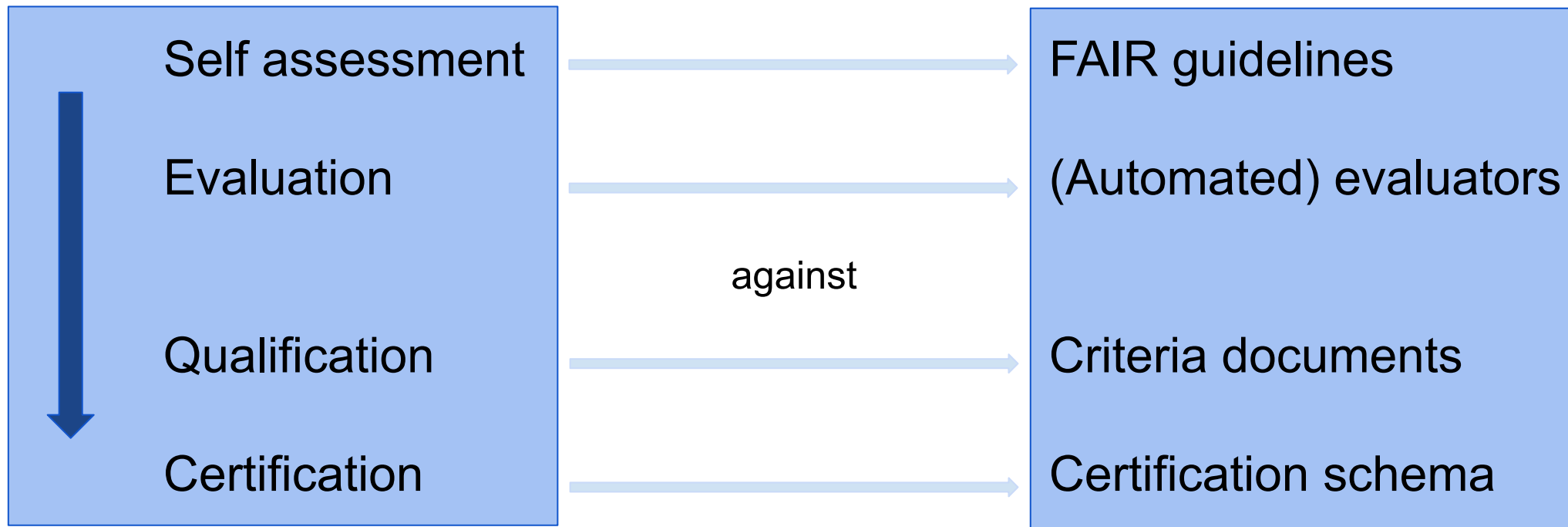
Where to start?

- Think what your repository needs
- You might wish to focus on:
 - FAIR (meta)data
 - CoreTrustSeal certification
 - certain aspects/ requirements of FAIR and/or CTS

Why are FAIR evaluations and certifications important?

- Increases “FAIR Awareness“ within organisations
- Increases the trust-factor of repositories and communities
- Preserves the FAIR Principles’ intent (Prevents “watering down”)
- Essential for interoperability within data-exchange projects
- Drives convergence across domains and countries

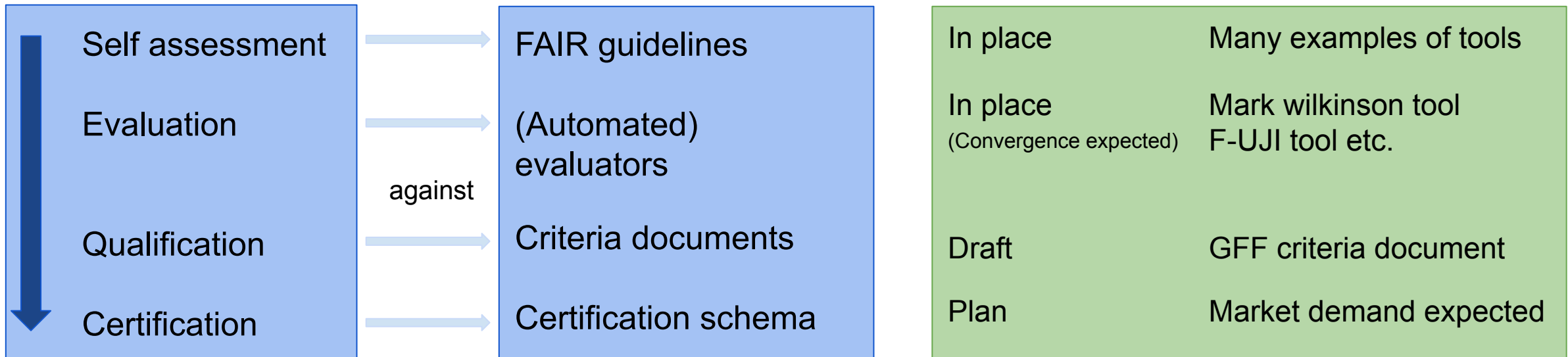
Qualification and certification is a step-by-step process



Parties involved in certification

Certification schema

- a. Schema development (Community supported initiatives)
- b. Schema holding (National or regional institutes)
- c. Schema execution (Accredited certifying bodies)

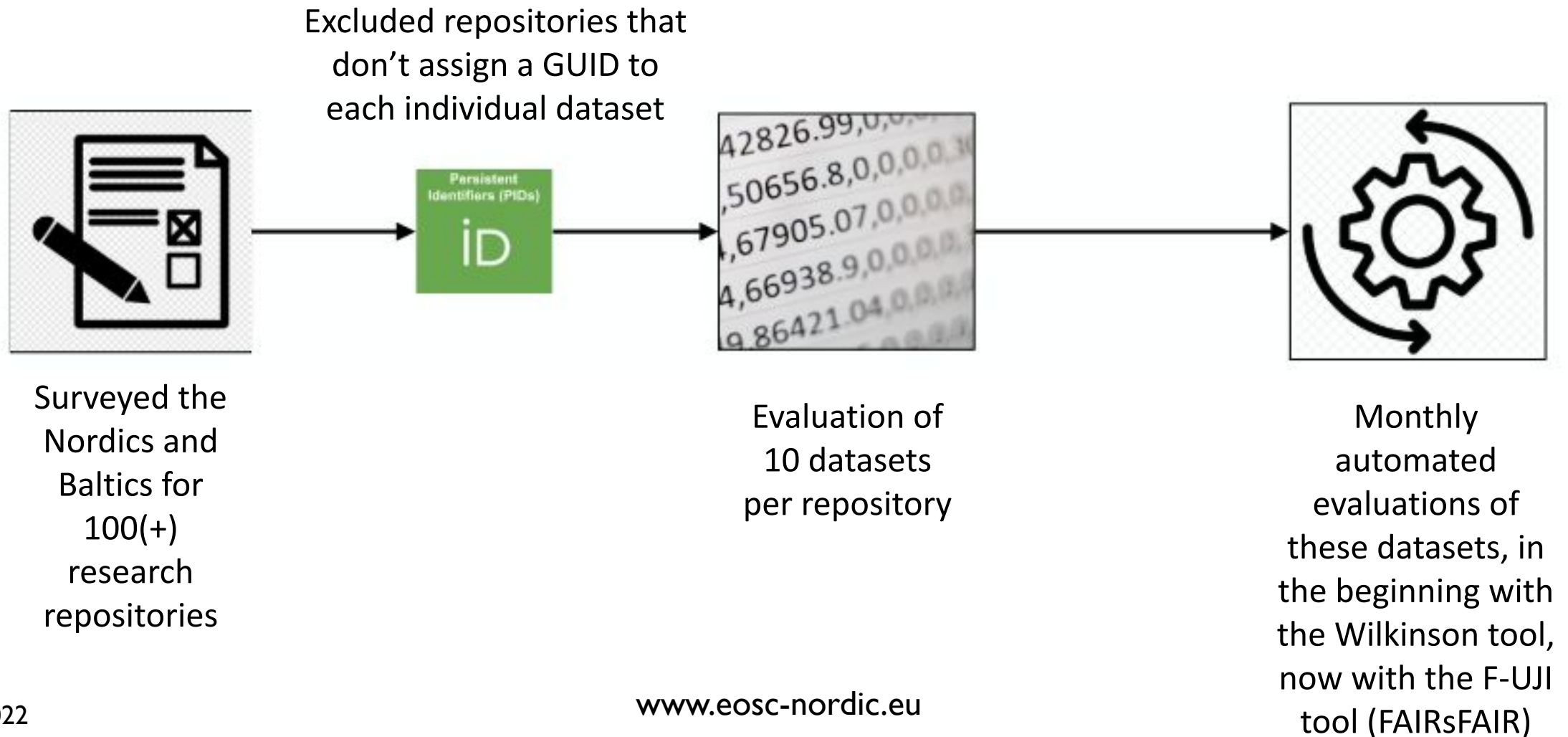


Things to consider

- FAIR awareness is more valuable than a high score
- Consider “Peer Review Process” for qualifications
- Articulate and Publish all FAIR implementation choices made, ideally in a FIP (FAIR Implementation Profile)
- Be aware that funders will continue to insist on Data Management Plans (DMP’s) for funding scientific research
- Focus on “Machine Actionability“ of data / metadata / DMPs

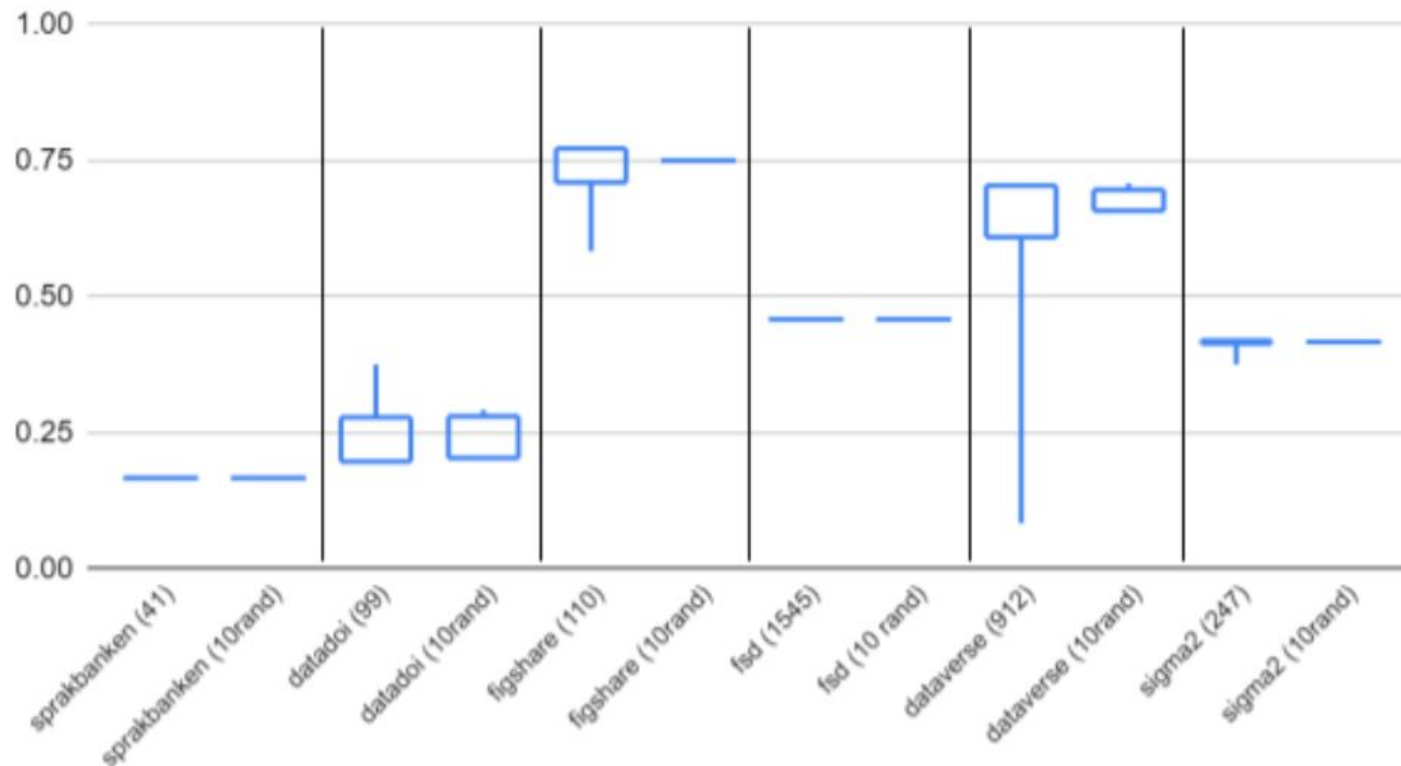
WP4: FAIR maturity of Nordic and Baltic Repositories

WP4 activities



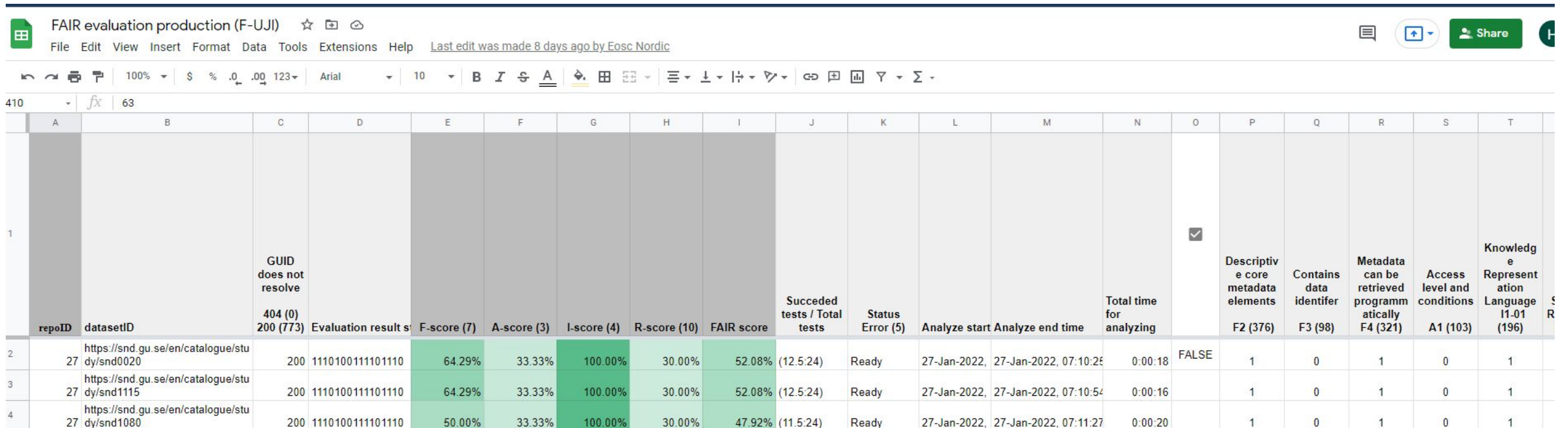
10 vs. all datasets

FAIR score



- Tested if 10 random selected datasets are statistically representative
- 6 repositories have OAI-PMH access, so we could extract all identifiers
- Little variation was found

The workflow



repoID	datasetID	GUID does not resolve 404 (0) 200 (773)	Evaluation results	F-score (7)	A-score (3)	I-score (4)	R-score (10)	FAIR score	Succeeded tests / Total tests	Status Error (5)	Analyze start	Analyze end time	Total time for analyzing	<input checked="" type="checkbox"/>	Descriptive core metadata elements F2 (376)	Contains data identifier F3 (98)	Metadata can be retrieved programmatically F4 (321)	Access level and conditions A1 (103)	Knowledge Representation Language I1-01 (196)
27	https://snd.gu.se/en/catalogue/study/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
27	https://snd.gu.se/en/catalogue/study/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
27	https://snd.gu.se/en/catalogue/study/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

The workflow

	F	A	I	R						
	38%	41%	37%	23%						
	(73)	(45)	(37)	(47)		72	72	72	72	72
	0.380	0.251	0.185	0.151	0.236	0.007	0.018	0.013	0.014	0.010
850	(73)	(73)	(73)	(73)						
								74%	<33%	
MAX	100.0%	100.0%	75.0%	0.5(3)	70.8%			15%	33%<X<50%	
MIN	14.3%	0.0%	0.0%	0.0%	4.2%			13%	>50%	

- Summary for the entire sample is generated automatically
- More data-analysis needs manual work

If you want to evaluate a dataset manually...

<https://www.f-uji.net/index.php>

F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics

[Click here to assess a dataset](#)

F-UJI was developed by Anusuriya Devaraju & Robert Huber (DANGAEA) under the umbrella of the FAIRsFAIR project.

FAIR assessment

F-UJI is a web service to programmatically assess FAIRness of research data objects (aka data sets) based on metrics developed by the FAIRsFAIR project.

Please use the form below to enter an identifier (e.g. DOI, URL) of the data set you wish to assess. Optionally you also can enter a metadata service (OAI-PMH, SPARQL, CSW) endpoint URI which F-UJI can use to identify additional information.

Research Data Object (URL/PID):*

Enter a valid PID or URL of the dataset's landing page (e.g. a DOI)

(Optional) Metadata service endpoint: (Optional) endpoint serving your metadata in different formats

Metadata service type: OAI-PMH

Use DataCite? ⓘ

[Start FAIR Assessment](#)

Assessment Results:

Evaluated Resource:

Meteorological site-level data from the CLIMAITE project at Brandbjerg, Denmark, from 1 October 2005 to 31 December 2013 V.3

[Save](#) [JSON](#) [New](#)

FAIR level: ⓘ **moderate**

Resource PID/URL: <https://doi.org/10.17894/ucph.e58a99c2-da7b-444a-b1c0-11f00e70041c>

DataCite support: enabled

Metric Version: metrics_v0.4

Metric Specification: <https://doi.org/10.5281/zenodo.4081213>

Software version: 1.4.6

Download assessment results: [JSON](#)

Save and share assessment results:

Summary:

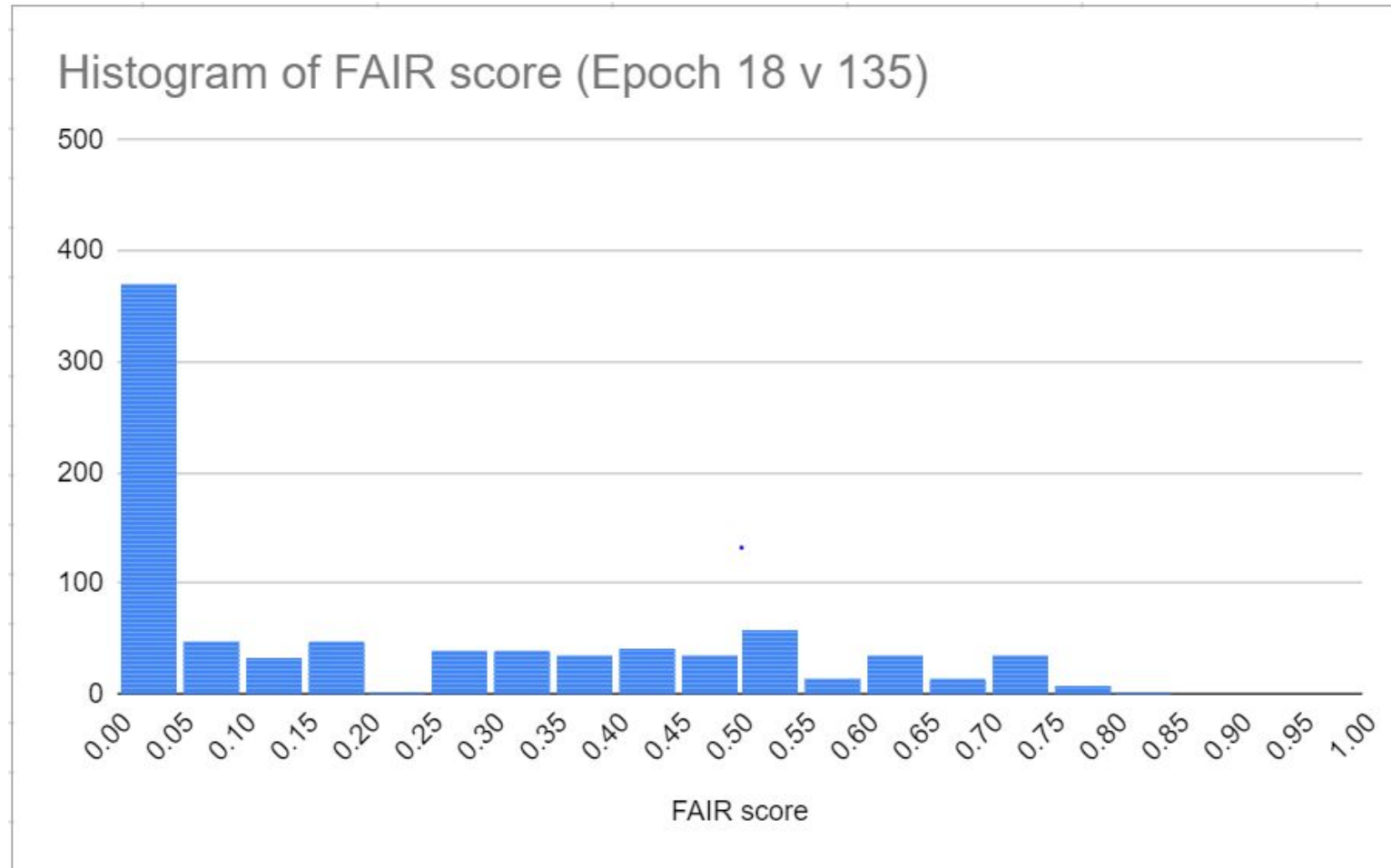
	Score earned:	Fair level:
Findable:	6 of 7	moderate
Accessible:	1 of 3	initial
Interoperable:	2 of 4	moderate
Reusable:	2 of 10	initial

Report:

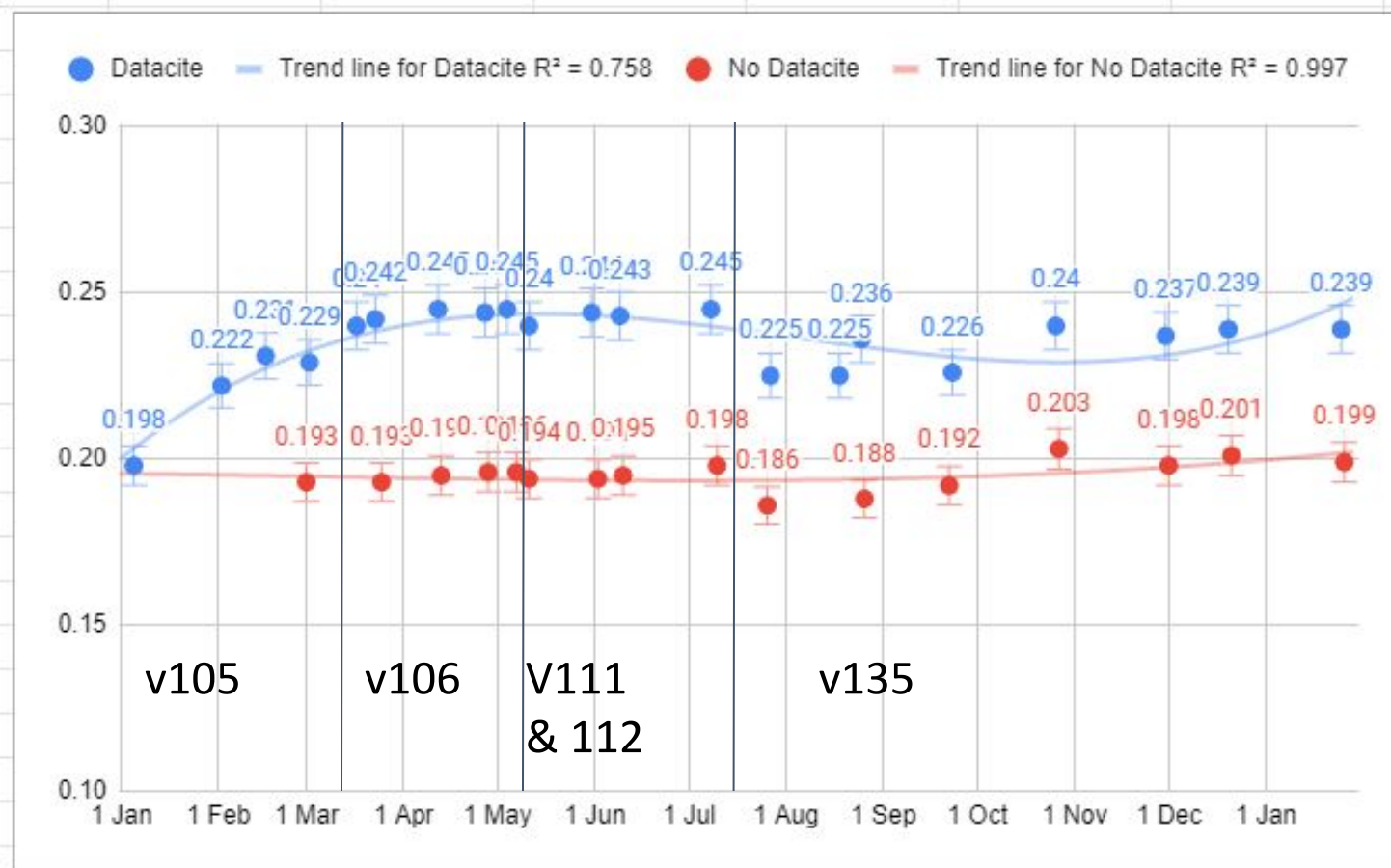
Findable

FsF-F1-01D - Data is assigned a globally unique identifier.	✓
FsF-F1-02D - Data is assigned a persistent identifier.	✓
FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.	✓
FsF-F3-01M - Metadata includes the identifier of the data it describes.	?

Histogram of FAIR scores of all evaluated repositories*

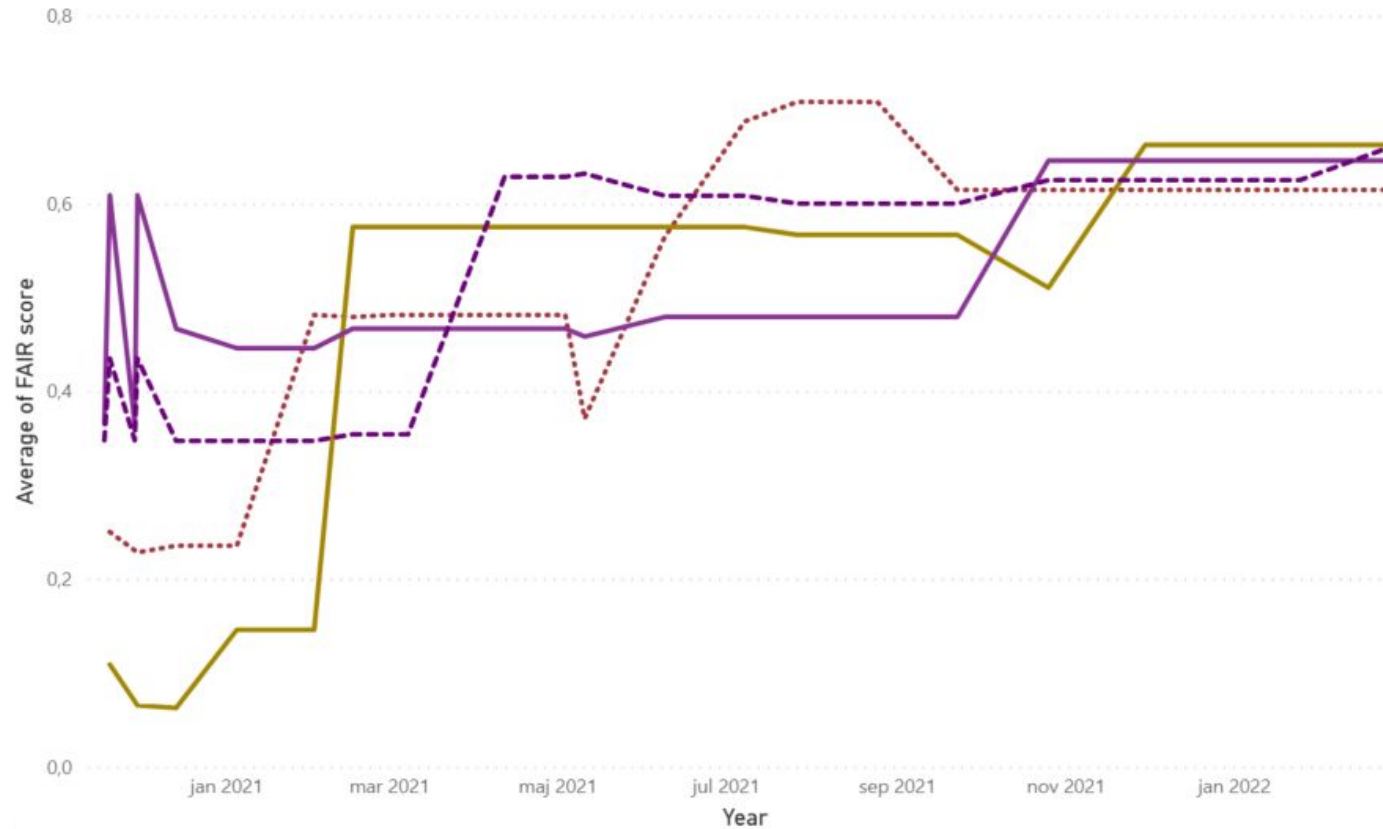


Preliminary results



- 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

Repositories with notable changes



- Evolution of FAIR score of selected repositories over time.
- repositories have been contacted
- relatively simple changes have been implemented

Solid purple: SND, dashed purple: Bolin Centre Database, yellow: ICOS, dashed red: QsarDB.

Lessons learnt

DO

- focus on metadata
 - Metadata available, but only some datasets can be downloaded without registration → FAIRness of metadata is crucial
- take basic steps:
 - Embedded JSON → multilingual and vocabulary based
 - Enriched DublinCore
 - Typed links / signposting
 - Vocabularies, ontologies, keywords, mappings...

DON'T

- do it for the evaluator
- worry if not reaching 10/10 : understand the results and limitations
- think FAIR only now. Keeping data FAIR needs to be addressed

Thank you!



www.eosc-nordic.eu

 https://twitter.com/EOSC_Nordic

 <https://www.linkedin.com/groups/13756550/>

 <https://www.linkedin.com/groups/13756550/>

A Journey from 0 to 75% Is that FAIR enough?

Guðbjörg Andrea Jónsdóttir

Director, DATICE and Social Science Research Institute

Webinar on User Experience with FAIR Evaluation Tools and Services
11th of October 2022

DATICE – The Icelandic Social Science Data Service

- Formally established in late 2018
- Funded by University of Iceland's School of Social Sciences
- National Service Provider for CESSDA, since 17th of June 2020
- Accepts social science data from all disciplines (e.g., survey data, panel data), as long as it's of traditional scale
- Has mainly archived and published data from political science, sociology, and ethnology (will accommodate data from other disciplines in the future)

DATICE's mission is to make research data findable, accessible, interoperable and reusable in accordance with international standards such as the FAIR principles. Our main objectives are to maximise the use of research data in Iceland, to ensure free and wide access to it and secure preservation of data for the long-term.



The FAIR principles



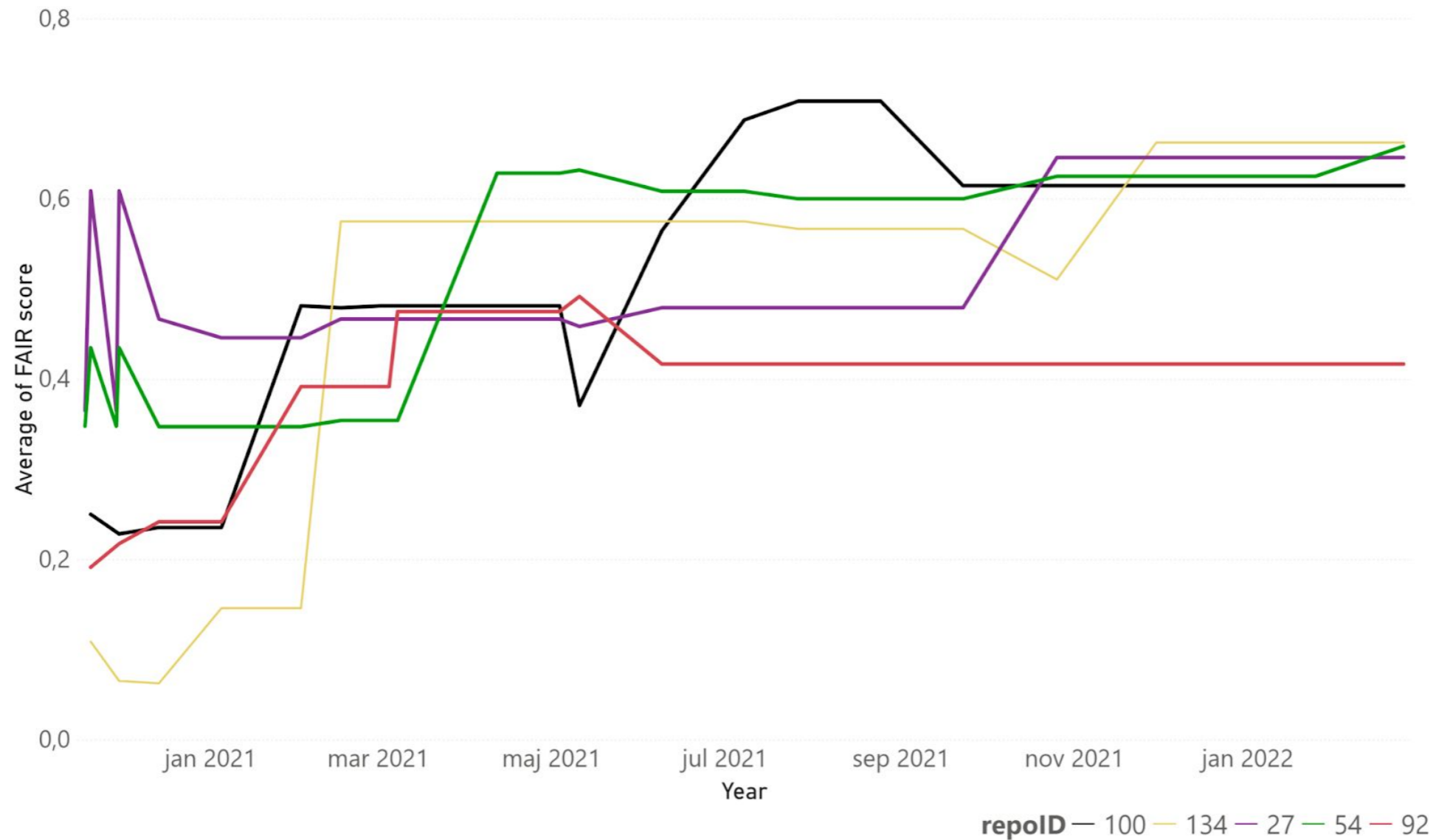
Persistent Identifier (PID)



GESIS Leibniz-Institut in Germany

da|ra
Registration agency for
social and economic data

Repositories with notable changes – EOSC-Nordic



F-UJI Evaluation – Daticce: November 2020 and April 2022

repoID	datasetID	F-score (7)	A-score (3)	I-score (4)	R-score (9)	FAIR score	Succeeded tests / Total tests	Status	Analyze start time
92	http://fel.hi.is/ISKOS1983	14.29%	33.33%	0.00%	0.00%	8.70%	(2:23)	Ready	20-Nov-2020, 14:42:46
92	http://fel.hi.is/ISKOS2003	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:42:58
92	http://fel.hi.is/ISKOS2017	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:43:09
92	http://fel.hi.is/ICENES1991	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:43:31
92	http://fel.hi.is/ICENES1999	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:45:33
92	http://fel.hi.is/ICENES2013	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:45:48
92	to_mental_health	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:46:00
92	http://fel.hi.is/a_study_of_luck_and_superstition_in_iceland	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:46:11
92	http://fel.hi.is/cultural_scales	14.29%	0.00%	0.00%	0.00%	4.35%	(1:23)	Ready	20-Nov-2020, 14:46:22
92	10.34881/1.00001	85.71%	66.67%	0.00%	33.33%	47.83%	(11:23)	Ready	20-Nov-2020, 14:59:22
92	10.34881/1.00006	71.43%	33.33%	0.00%	33.33%	47.83%	(9:23)	Ready	20-Nov-2020, 14:59:37
92	10.34881/1.00011	71.43%	33.33%	0.00%	33.33%	39.13%	(9:23)	Ready	20-Nov-2020, 14:59:49
92	10.34881/1.00003	71.43%	33.33%	0.00%	33.33%	39.13%	(9:23)	Ready	20-Nov-2020, 15:00:01
92	10.34881/1.00005	71.43%	33.33%	0.00%	33.33%	39.13%	(9:23)	Ready	20-Nov-2020, 15:00:13
92	10.34881/1.00009	71.43%	33.33%	0.00%	33.33%	39.13%	(9:23)	Ready	20-Nov-2020, 15:02:17
92	10.34881/1.00001	92.86%	33.33%	75.00%	40.00%	60.42%	(14.5:24)	Ready	27-Apr-2022, 11:40:02
92	10.34881/1.00006	92.86%	33.33%	75.00%	40.00%	60.42%	(14.5:24)	Ready	27-Apr-2022, 11:40:32
92	10.34881/1.00011	92.86%	33.33%	75.00%	40.00%	60.42%	(14.5:24)	Ready	27-Apr-2022, 11:41:00
92	10.34881/1.00003	92.86%	33.33%	75.00%	40.00%	60.42%	(14.5:24)	Ready	27-Apr-2022, 11:41:31
92	10.34881/1.00005	92.86%	33.33%	75.00%	40.00%	60.42%	(14.5:24)	Ready	27-Apr-2022, 11:41:54
92	10.34881/1.00009	92.86%	33.33%	75.00%	40.00%	60.42%	(14.5:24)	Ready	27-Apr-2022, 11:42:23
92	to_mental_health	14.29%	0.00%	0.00%	0.00%	4.17%	(1:24)	Ready	27-Apr-2022, 11:42:45
92	http://fel.hi.is/a_study_of_luck_and_superstition_in_iceland	14.29%	0.00%	0.00%	0.00%	4.17%	(1:24)	Ready	27-Apr-2022, 11:42:58
92	http://fel.hi.is/cultural_scales	14.29%	0.00%	0.00%	0.00%	4.17%	(1:24)	Ready	27-Apr-2022, 11:43:18



Íslensk gagnasöfn

► Fordómar í alþjóðlegu samhengi

Íslenska kosningarannsóknin

1983

1987

1991

1995

1999

2003

2007

2009

2013

2016

2017

Íslenska

Íslenska kosningarannsóknin 2017

Gagnaskrá og tilheyrandi skjöl eru að finna hér að neðan. Að auki er hlekkur á gagnvirka greiningu á netinu, í NESSTAR WebView, þar sem auðvelt er að skoða lýsandi tölfræði o.fl.

DOI númer	10.34881/1.00011
Útgáfa gagnaskrár	1.0.0
Höfundur/höfundar	<ul style="list-style-type: none">Önnudóttir, Eva Heiða (Stjórn málafræðideild, Háskóli Íslands)Harðarson, Ólafur Þórður (Stjórn málafræðideild, Háskóli Íslands)Þórisdóttir, Hulda (Stjórn málafræðideild, Háskóli Íslands)Helgason, Agnar Freyr (Stjórn málafræðideild, Háskóli Íslands)
Útgáfudagur	2020-01-31
Umsjón gagnasöfnunar	Félagsvísindastofnun, Háskóli Íslands
Fjármögnun	<ul style="list-style-type: none">Althingi (the Icelandic Parliament)Rannsóknasjóður Háskóla Íslands (University of Iceland Research Fund)

DATICE is the official service provider for the Consortium of European Social Science Data Archives (CESSDA ERIC) in Iceland and is a collector and supplier of high-quality research data. The service is located within the Social Science Research Institute (SSRI) at University of Iceland.



School of Social Science / Social
Science Research Institute



School of Education / Education
Research Institute



Icelandic National Election Study
(ICENES)



COVID-19 Studies in the Social
and Education Sciences

Search this dataverse...

Find

Advanced Search

Datasets (4)

Files (24)

Files (124)

Dataverse Category

Research Project (2)

Department (1)

Organization or Institution (1)

Publication Year

2022 (2)

2021 (22)

2020 (3)

Keyword Term

Voting behavior (11)

Elections (10)

Political attitudes (10)

Parliamentary elections (9)

Voting (9)

More...

1 to 10 of 28 Results

Sort

Health Behaviour of Icelandic Youth

Jan 26, 2022

Jóhannsson, Erlingur Sigurður; Kristjánsdóttir, Ása Guðrún; Hrafnkelsson, Hannes; Magnússon, Kristján Þór; Stefánsdóttir, Rúna Sif; Guðmundsdóttir, Sigríður; Arngrímsson, Sigurbjörn Árni; Hrafnkelsdóttir, Soffía; Rögnvaldsdóttir, Vaka, 2022, "Health Behaviour of Icelandic Youth", <https://doi.org/10.34881/ZYG5XM>, GAGNÍS (DATICE), V1

Health Behavior of Icelandic Youth is an extensive long-term study on the health status of Icelandic youth and its relation to sleep, physical activity and school environment. The research is organized by investigators at the School of Education, University of Iceland, in collabo...

School of Education / Education Research Institute (University of Iceland)

Jan 5, 2022



Fordómar í alþjóðlegu samhengi: Rannsókn á viðhorfum Íslendinga til geðrænna vandamála

Dec 9, 2021

Bernburg, Jón Gunnar; Ólafsdóttir, Sigrún; Þórlindsson, Þórólfur; Pescosolido, Bernice A., 2021, "Fordómar í alþjóðlegu samhengi: Rannsókn á viðhorfum Íslendinga til geðrænna vandamála", <https://doi.org/10.34881/1.00021>, GAGNÍS (DATICE), V2, UNF:6:YclqbdNB9Kgj+eaeQd4MA== [fileUNF]

Fordómar í alþjóðlegu samhengi: Rannsókn á viðhorfum Íslendinga til geðrænna vandamála, er hluti af Stigma in Global Context: Mental Health

F-UJI evaluation

Assessment Results (saved at 2022-05-25 21:06:28):

Evaluated Resource:

Íslenska kosningarannsóknin 2017

[↓ {JSON}](#)

[↻ New](#)

FAIR level: [?](#)

moderate

Resource PID/URL:

10.34881/1.00011

DataCite support:

enabled

Metric Version:

metrics_v0.4

Metric Specification:

<https://doi.org/10.5281/zenodo.4081213>

Software version:

1.4.9b

Download saved assessment results:

[{JSON}](#)

Save and share assessment results:

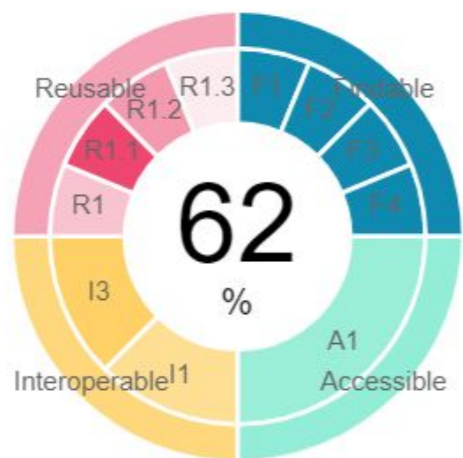
Saved assessments:

- [2022-05-25](#)  **moderate**
- [2022-10-05](#)  **moderate**

F-UJI evaluation – ICENES 2017

May 2022

Summary:



	Score earned:		Fair level:
Findable:	7 of 7	<input type="radio"/>	advanced
Accessible:	1 of 3	<input type="radio"/>	initial
Interoperable:	3 of 4	<input type="radio"/>	moderate
Reusable:	4 of 10	<input type="radio"/>	initial

FsF-A1-03D - Data is accessible through a standardized communication protocol.



FAIR level: 0 of 3

incomplete

Score: 0 of 1

Output:
{
 "standard_data_protocol": null
}

Metric tests:	Test:	Test name:	Score:	Maturity:	Result:
	FsF-A1-03D-1	Metadata includes a resolvable link to data based on standardized web communication protocols.			

Debug messages:

Level:	Message:
INFO	NO content (data) identifier is given in metadata

FsF-R1.3-02D - Data is available in a file format recommended by the target research community.



FAIR level: 0 of 3

incomplete

Score: 0 of 1

Output: []

Metric tests:

Test:	Test name:	Score:	Maturity:	Result:
FsF-R1.3-02D-1	The format of a data file given in the metadata is listed in the long term file formats, open file formats or scientific file formats controlled list			
a	The format of the data file is an open format			
b	The format of the data file is a long term format			
c	The format of the data file is a scientific format			

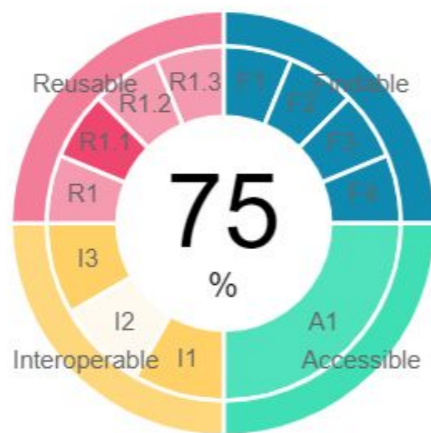
Debug messages:

Level:	Message:
INFO	File format(s) specified -: ['application/pdf', 'text/tab-separated-values']
WARNING	Could not perform file format checks as data content identifier(s) unavailable/inaccessible

F-UJI evaluation – ICENES 2017

October 2022

Summary:



	Score earned:		Fair level:
Findable:	7 of 7	○	advanced
Accessible:	2 of 3	○	moderate
Interoperable:	3 of 4	○	moderate
Reusable:	6 of 10	○	moderate

FsF-A1-03D - Data is accessible through a standardized communication protocol.



FAIR level: 3 of 3

advanced

Score: 1 of 1

Output:

```
{
  "standard_data_protocol": {
    "https": "Hyper Text Transfer Protocol Secure"
  }
}
```

Metric tests:	Test:	Test name:	Score:	Maturity:	Result:
	FsF-A1-03D-1	Metadata includes a resolvable link to data based on standardized web communication protocols.	1	3	

Debug messages:

Level:	Message:
SUCCESS	Standard protocol for access to data object found: https



FAIR level: 2 of 3

moderate

Score: 1 of 1

Output:

```
[
  {
    "file_uri": "https://dataverse.rhi.hi.is/api/access/datafile/198",
    "is_preferred_format": true,
    "is_recommended_format": true
  }
]
```

Metric tests:

Test:	Test name:	Score:	Maturity:	Result:
FsF-R1.3-02D-1	The format of a data file given in the metadata is listed in the long term file formats, open file formats or scientific file formats controlled list	1		
a	The format of the data file is an open format		1	
b	The format of the data file is a long term format		2	
c	The format of the data file is a scientific format			

Debug messages:

Level:	Message:
INFO	Data content identifier provided -: ['https://dataverse.rhi.hi.is/api/access/datafile/199', 'https://dataverse.rhi.hi.is/api/access/datafile/196', 'https://dataverse.rhi.hi.is/api/access/datafile/198', 'https://dataverse.rhi.hi.is/api/access/datafile/195', 'https://dataverse.rhi.hi.is/api/access/datafile/201']
SUCCESS	Could identify a file format commonly used by the scientific community -:text/tab-separated-values

DATICE

The Icelandic Social Science Data Service

Using F-UJI to assess FAIRness of CESSDA Data Catalogue

*John Shepherdson, Matthew Morris / CESSDA MO
Kostas Papagiannopoulos / EKKE*

11 October 2022 / MDO webinar

 cessda.eu

 @CESSDA_Data

Using F-UJI to assess FAIRness of CDC

Problem statement

Automating F-UJI assessments

Improving the FAIRness of CDC

Disseminating the results

Issues and next steps

Problem statement

Need to assess the FAIRness of the CESSDA Data Catalogue
Overview <https://www.cessda.eu/Tools/Data-Catalogue>

Contains more than 42,500 unique metadata records
55,000 plus research objects to assess (as some available in multiple languages)

Not really practical to manually assess 1% sample with [online F-UJI tool](#)

Need to automate the assessment process

Automating F-UJI assessments

Download and deploy the containerised [F-UJI API](#)

Runs in CESSDA's cloud-based infrastructure

Kubernetes clusters, Docker containers, Helm deployment charts

Create a helper application to call the API against each record in CDC in turn

Java code, runs in CESSDA's cloud-based infrastructure

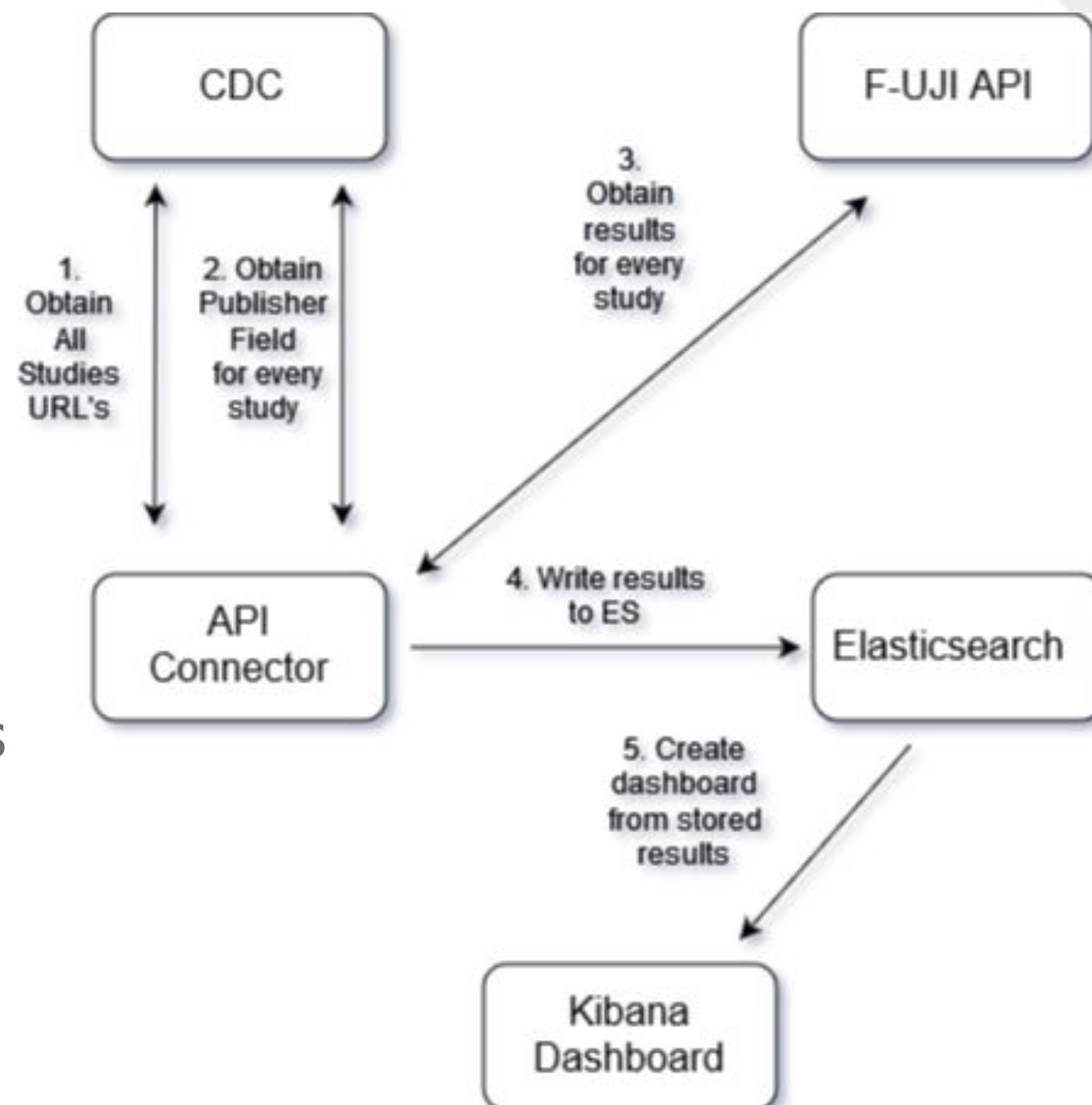
Iterates through CDC sitemap

Automating F-UJI assessments (2)

Store and index the FAIR assessments
Create a JSON file for each assessment result, store in Cloud storage bucket
Generate ElasticSearch index

Create a dashboard and use it to display the results at various levels of detail

Needed to add Publisher details to JSON files



Improving the FAIRness of CDC

Iterative approach

CDC 'as is' scored zero as JSON-LD not found

Was generated on demand using Javascript, hence invisible to F-UJI

CDC with static HTML pages scored 30% on average

Server-side JSON-LD is visible to F-UJI

Improving the FAIRness of CDC

CDC with improved DOI presentation scored > 50% on average

From e.g. *"pid": "http://doi.org/10.11587/0BVRTM (DOI), MZ9703 (WISDOM number)"*

To *"http://doi.org/10.11587/0BVRTM"*

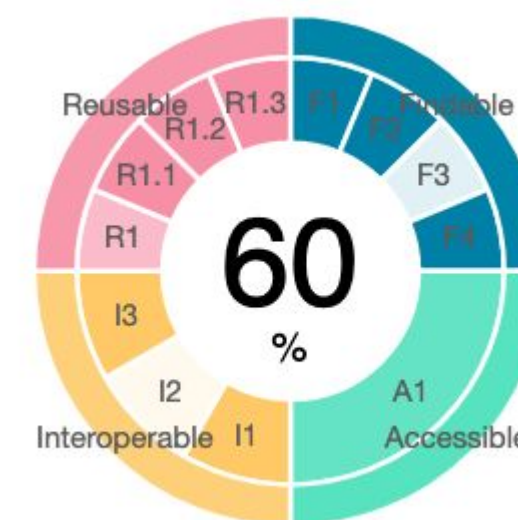
Records may contain multiple PIs, so priority order set to:

DOI > Handle > URN > ARK

CDC with signposting scored 60%

Link to external machine-readable metadata using HTTP headers (CDC

OAI-PMH endpoint: <https://datacatalogue.cessda.eu/oai-pmh/v0/oai>)



Disseminating the results

Using Kibana dashboards linked to Elasticsearch index

Summary - total no of records, number of passes/number of fails

Per Publisher - no of records, number of passes/number of fails

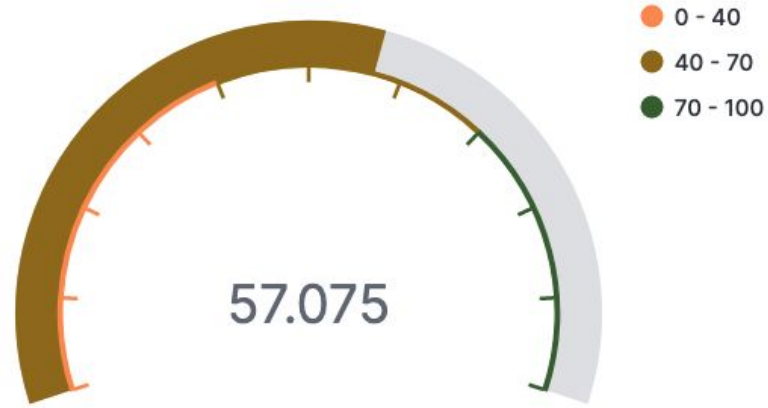
Per Publisher - list of failed records

Per Record - link to F-UJI online assessment (so we don't have to display all diagnostic data)

Need to map from F-UJI numeric scores to incomplete (0), initial (1) , moderate (2) and advanced (3)

CDC FAIR assessment dashboard

uji-cessda-overall-percent-score-new



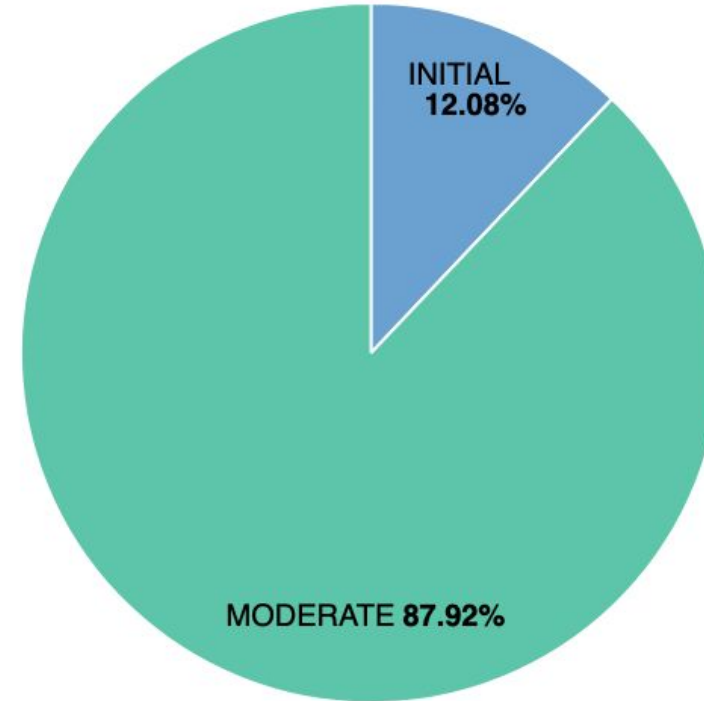
Average fuji-percent-score

RECORDS NUMBER

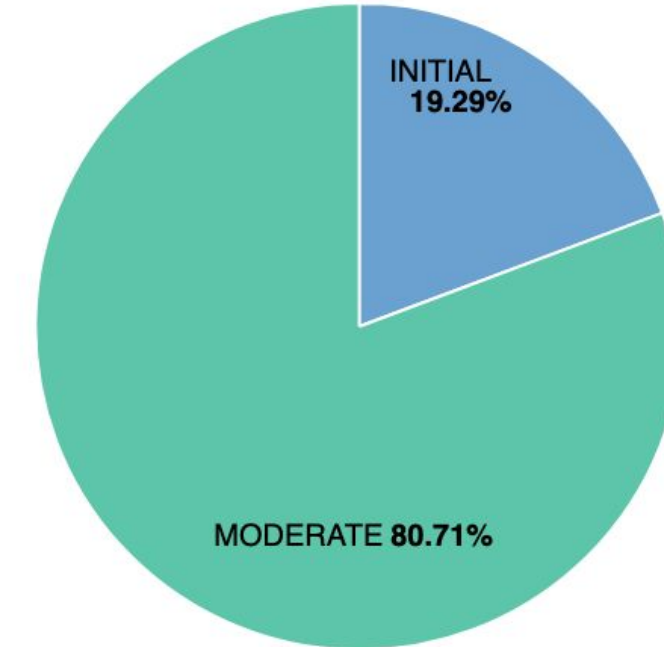
49,877

Number of
Records

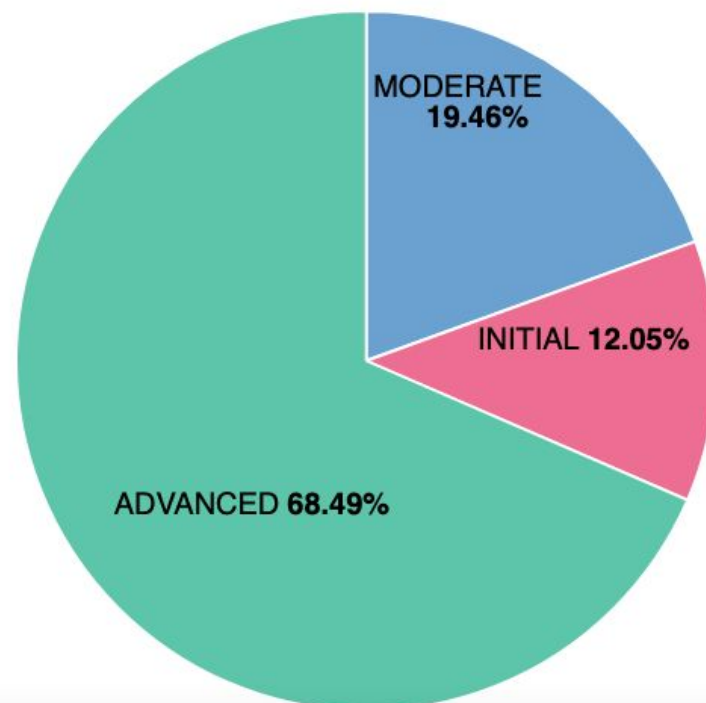
OVERALL FAIR RATING



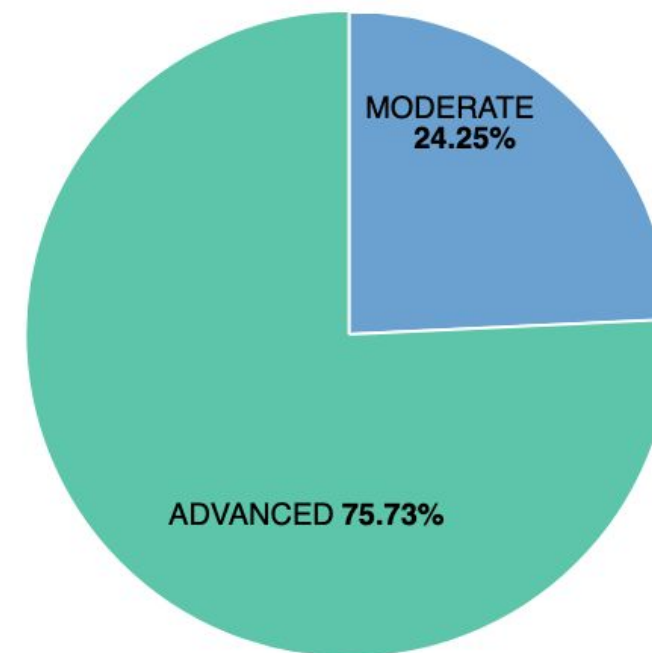
OVERALL ACCESSIBILITY RATING



OVERALL FINDABILITY RATING



OVERALL INTEROPERABILITY RATING



Issues and next steps

F-UJI API rate limiting (max 100 calls per minute)

Need to gap the calls to the API

Currently takes approx 55 hours to assess all CDC records

Cannot access URLs that use basic authorisation

Cannot test CDC FAIRness improvements before release to production

Being addressed by F-UJI developers

Cannot run the F-UJI web application automatically

Still have to press the button manually

Had to add publisher details to each result (not found in F-UJI output)

Issues and next steps

Bulk assessment process is expensive
Currently we make 55k calls to the API
Could we make 1 call and pass in an array of URLs?

Further improvements to CDC FAIRness

Low scores mostly relate to the data associated with the metadata record

Following are required:

- file name, size, type, PID/URL for data download
- data access level and conditions
- standard web protocol for data access
- namespaces of known semantic resources
- licence information for data reuse
- data file formats that match controlled list

	Score earned:		Fair level:
Findable:	6 of 7	🔄	moderate
Accessible:	1.5 of 3	🔄	initial
Interoperable:	3 of 4	🔄	moderate
Reusable:	4 of 10	🔄	initial

Will look at each in turn, to see what gains can be made, if any

Thanks for listening

Any questions?

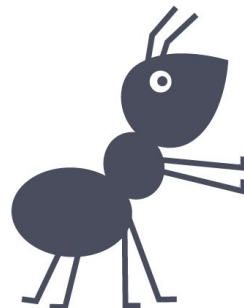
KonsortSWD



Consortium for the
Social, Behavioural, Educational
and Economic Sciences

CESSDA,
Webinar on User Experience
with FAIR Evaluation Tools and
Services

Presentation
11.10.2022



Application of 'RDA FAIR Data Maturity Model' to assess the PID registration service in terms of FAIRness

Janete Saldanha Bach
Claus-Peter Klas
Peter Mutschke

GESIS – Leibniz Institute for the Social Sciences

Janete Saldanha Bach



Janete Saldanha Bach is a Researcher at GESIS – Leibniz Institute for the Social Sciences, based in the Knowledge Technologies (KTS) Department, team FAIR Data and Human Information Interaction, working in the consortia KonsortSWD Project of the National Research Data Infrastructure (NFDI). She holds a Ph.D. and a Master's degree in Technology and Society interactions from the Federal and a bachelor's degree in Library Science. Her research expertise is in Open Science, especially in research data management and data reuse in the Social Sciences. She is currently involved in consortia KonsortSWD, Task Area 5 Measure 1 - developing the conceptual framework for the PID registration service at a variable level.

Claus-Peter Klas

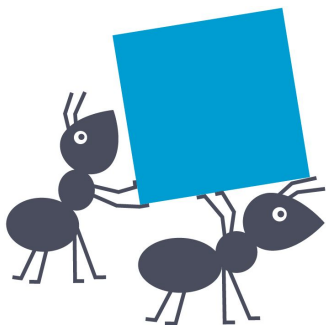


Claus-Peter Klas is lead of the Data & Service Engineering team in the department Knowledge Technologies for the Social Sciences of GESIS. He received his PhD in computer science at the University of Duisburg-Essen and was a postdoctoral researcher in the Department of Multimedia and Internet Applications, Faculty of Mathematics and Computer Science, University of Hagen, Germany. His research focuses on information retrieval, interactive information retrieval, information systems, databases, digital libraries, preservation and grid and cloud architectures. He developed the software Daffodil founded on a nation research project and worked in national and European research projects such as The European Film Gateway, SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg) and Smart Vortex (Scalable Semantic Product Data Stream Management for Collaboration and Decision Making in Engineering). He is currently responsible for several infrastructure projects within GESIS, such as daJra, SowiDataNet or Missy, all concerned with providing information and data for social scientists. In addition, he lead the measure PID Services in the national research infrastructure project NFDI. In his team, they are developing a open source DDI suite to support getting DDI into operation.

Peter Mutschke




Peter Mutschke is deputy head of the department "Knowledge Technologies for the Social Sciences (KTS)" and leader of the team "FAIR Data and Human Information Interaction" of KTS. His research interests include Information Retrieval, Network Analysis and Open Science. He worked in a number of national and international research projects, such as the DFG projects DAFFODIL and IRM and the EU projects WeGov, SENSE4US, OpenMinTeD and MOVING. Peter served as a member of the management committee of the Leibniz research alliance "Science 2.0/Open Science" from 2013-2021. He founded and coordinates the GO FAIR Implementation Network "Cross-Domain Interoperability of Heterogeneous Research Data (Go Inter)", and he is member of the steering committee of the FAIR Digital Objects Forum (fairdo.org) where he also co-chairs a working group on semantics. He is currently involved in consortia KonsortSWD, NFDI4DataScience and BERD@NFDI of the National Research Data Infrastructure (NFDI). ORCID: <https://orcid.org/0000-0003-3517-8071>.



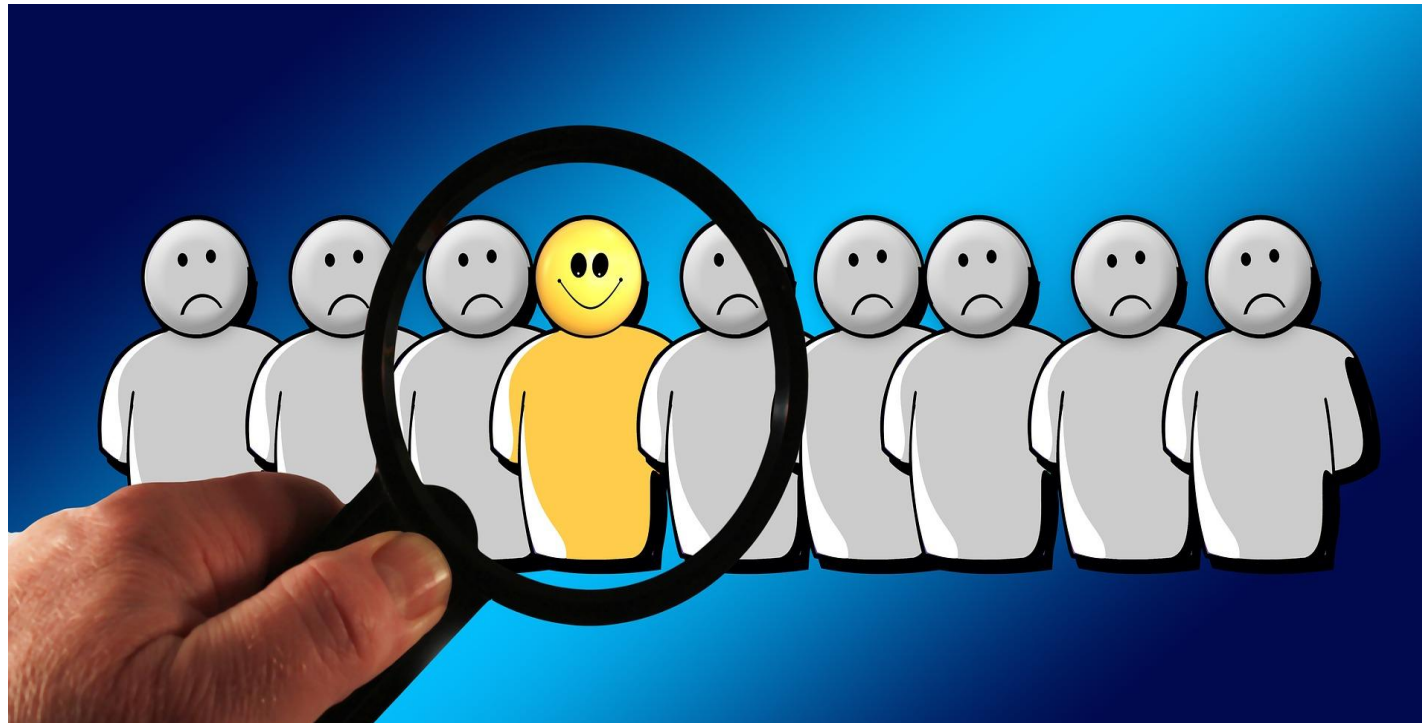
Agenda

- The PID Registration service
 - General goal and claim
 - The Research data granularity levels
 - Data citation using PIDs
- The PID Registration service: FAIR maturity level assessment
 - Criteria
 - Methodology
 - Results
 - Outcomes

- 
- General goal and claim

Claim

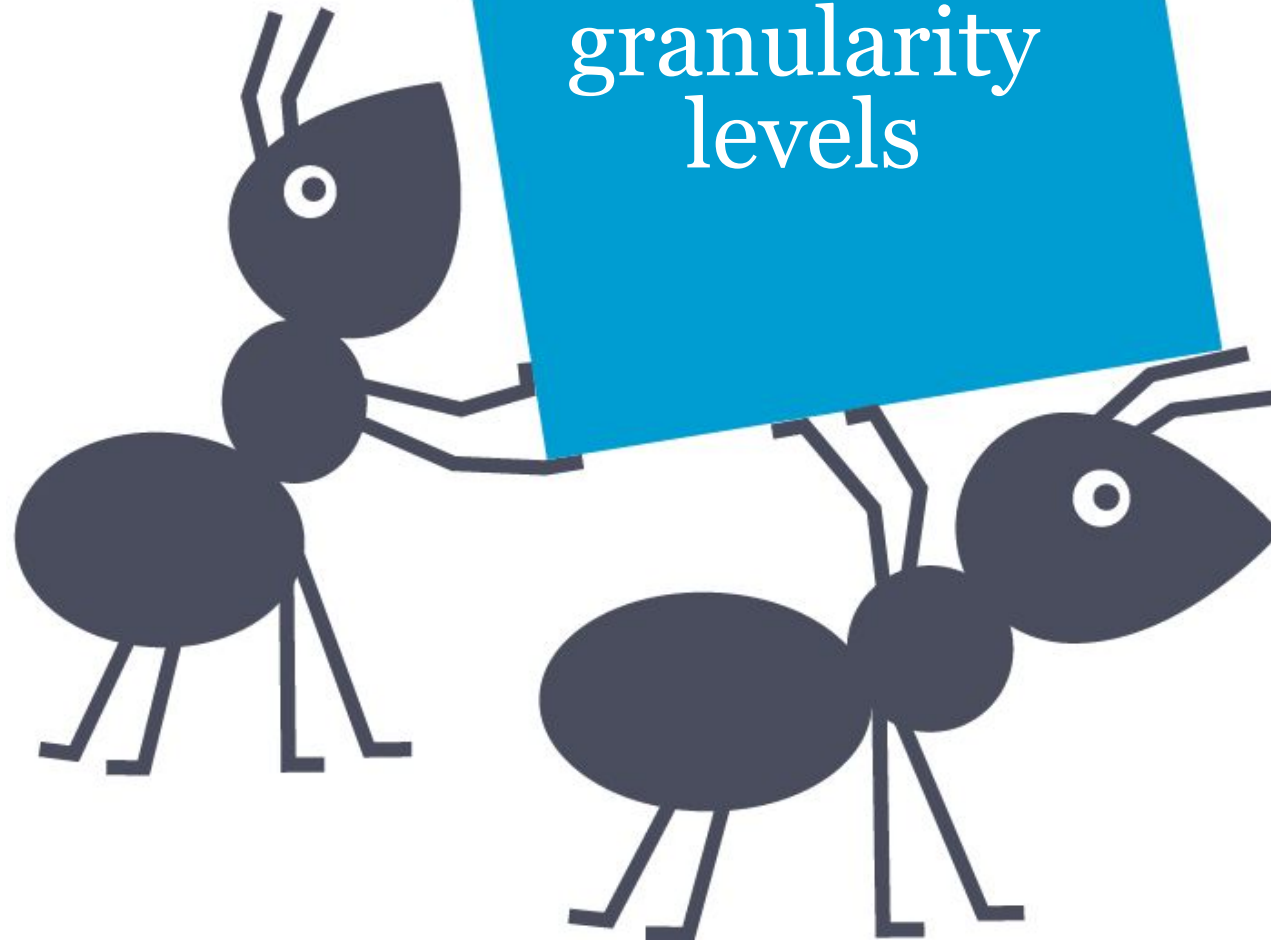
- Assigning a PID to a whole dataset is insufficient to unambiguously identify the information used and ensure an accurate data citation, thus, constraining the research results' trustworthiness.

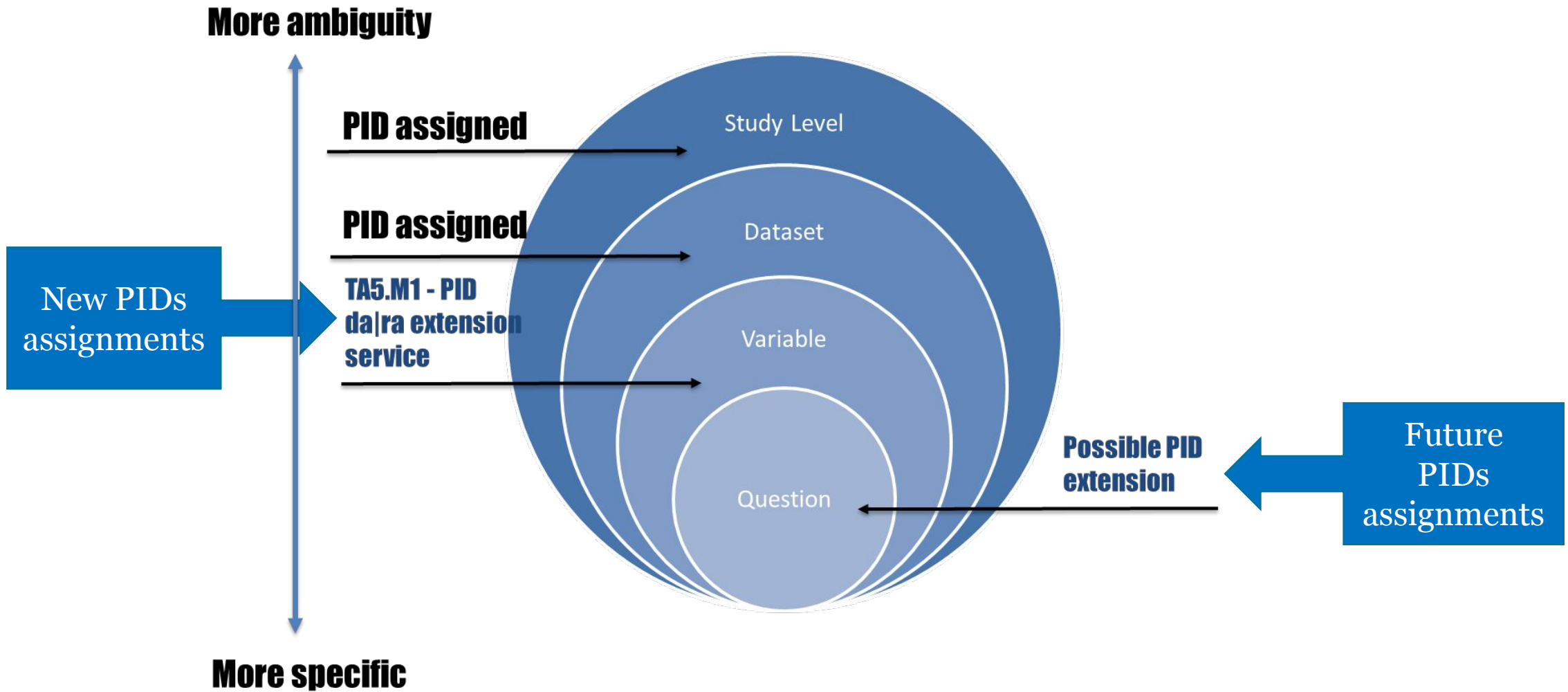


General goals

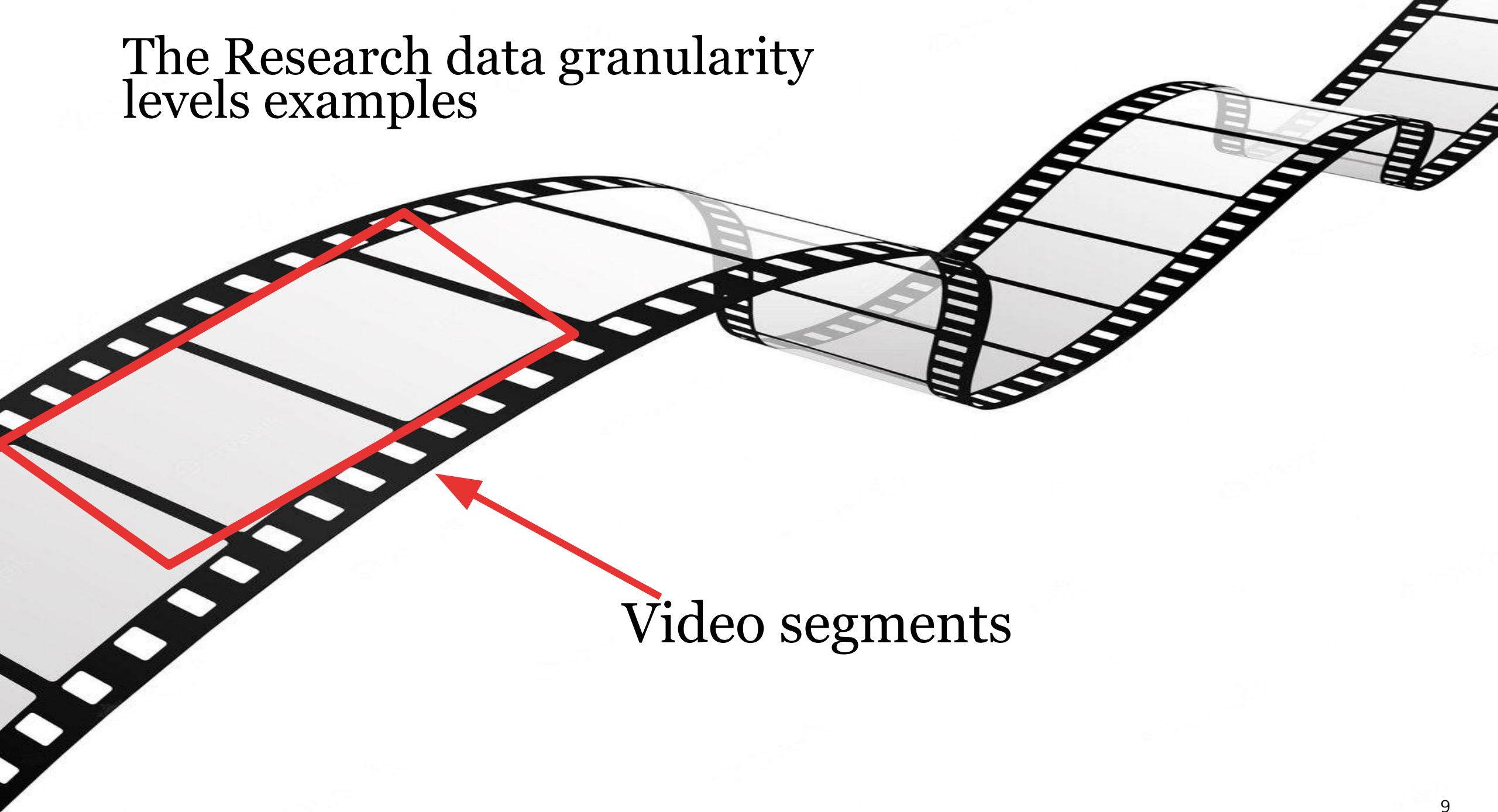
- Identify survey variables, using one identifier – the PID – will simplify FAIR data management to boost subsequent citation, get direct (meta)-data access, and data reuse
- Since PIDs are machine-actionable, they are used as technical bridges to the FAIR principles that can increase traceability and foster reproducibility of research results in the Social and Economic Sciences

The Research data granularity levels





The Research data granularity levels examples



Video segments

The Research data granularity levels examples

Use Cases

Questions

The image shows a woman holding a tablet that displays a survey questionnaire. The screen has a teal header with the text "Survey Questionnaire". Below the header, there are two main sections. The top section is titled "Question 1" and is highlighted with a red box. It contains three radio button options: "Yes", "No", and "Don't know". To the right of this section is a "Comments" field. The bottom section is titled "Question 2" and also contains a "Comments" field. A red arrow points from the word "Questions" above to the "Question 1" section on the tablet screen.

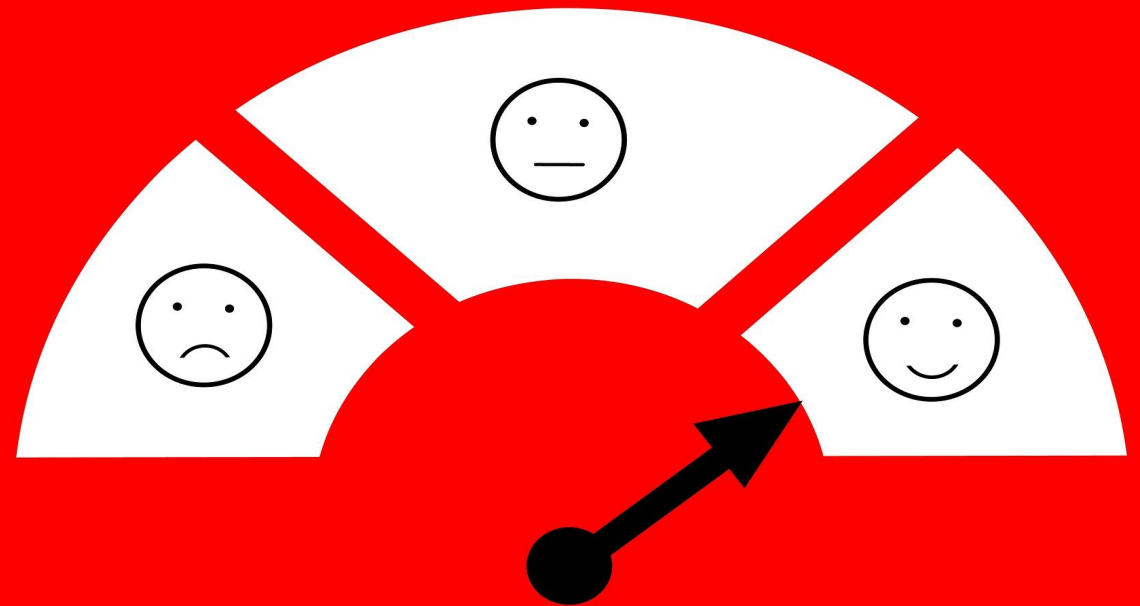
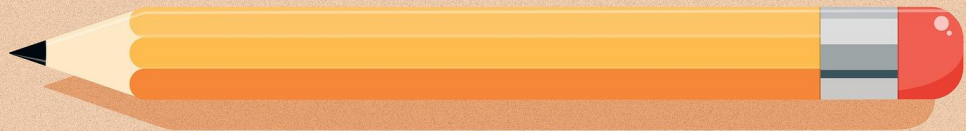
Question	Options	Comments
Question 1	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know	
Question 2	<input type="checkbox"/> Yes	

Measures Scales



Yes

No

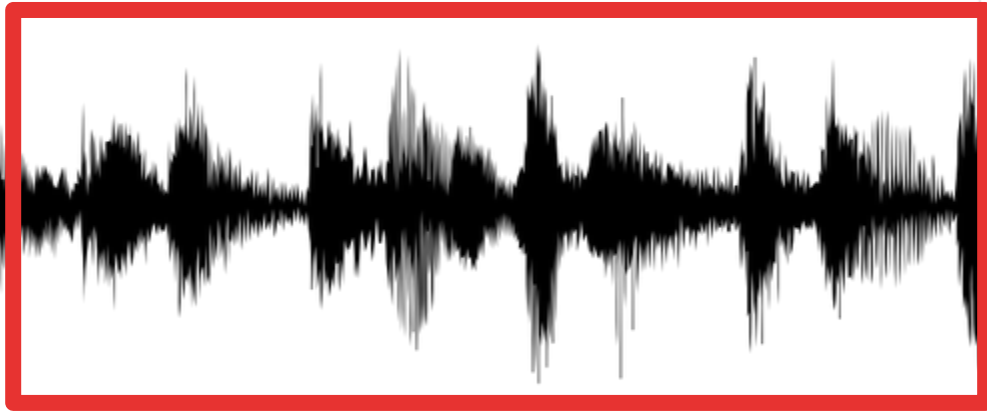


The Research data granularity levels examples

Audio segments



Audio files

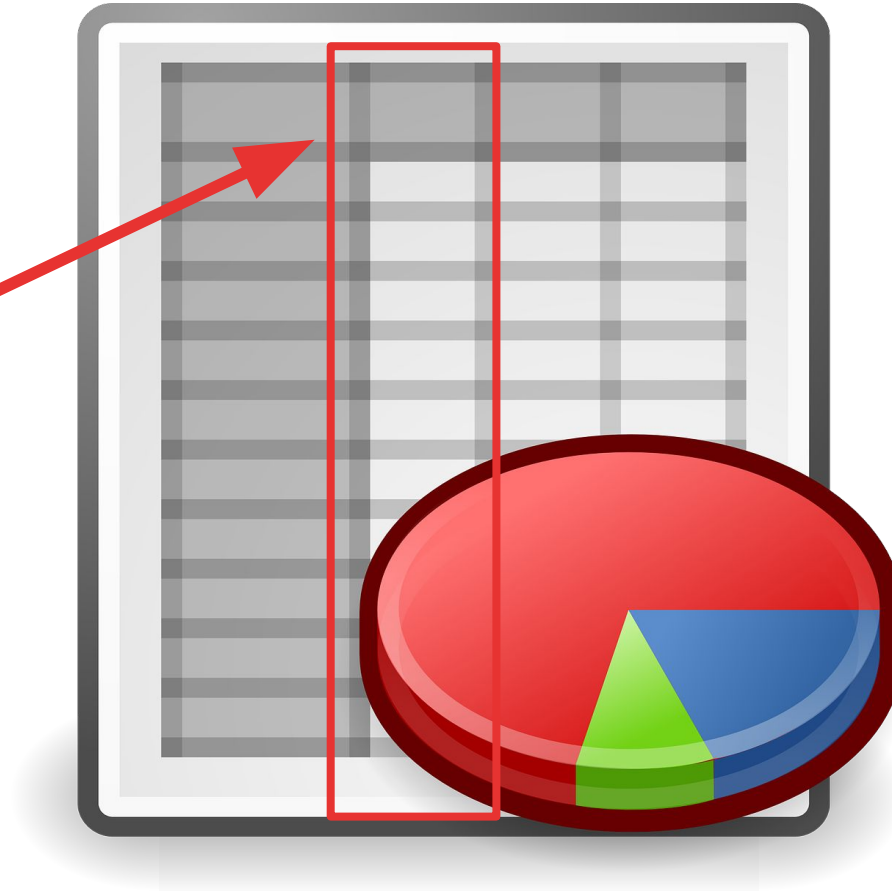


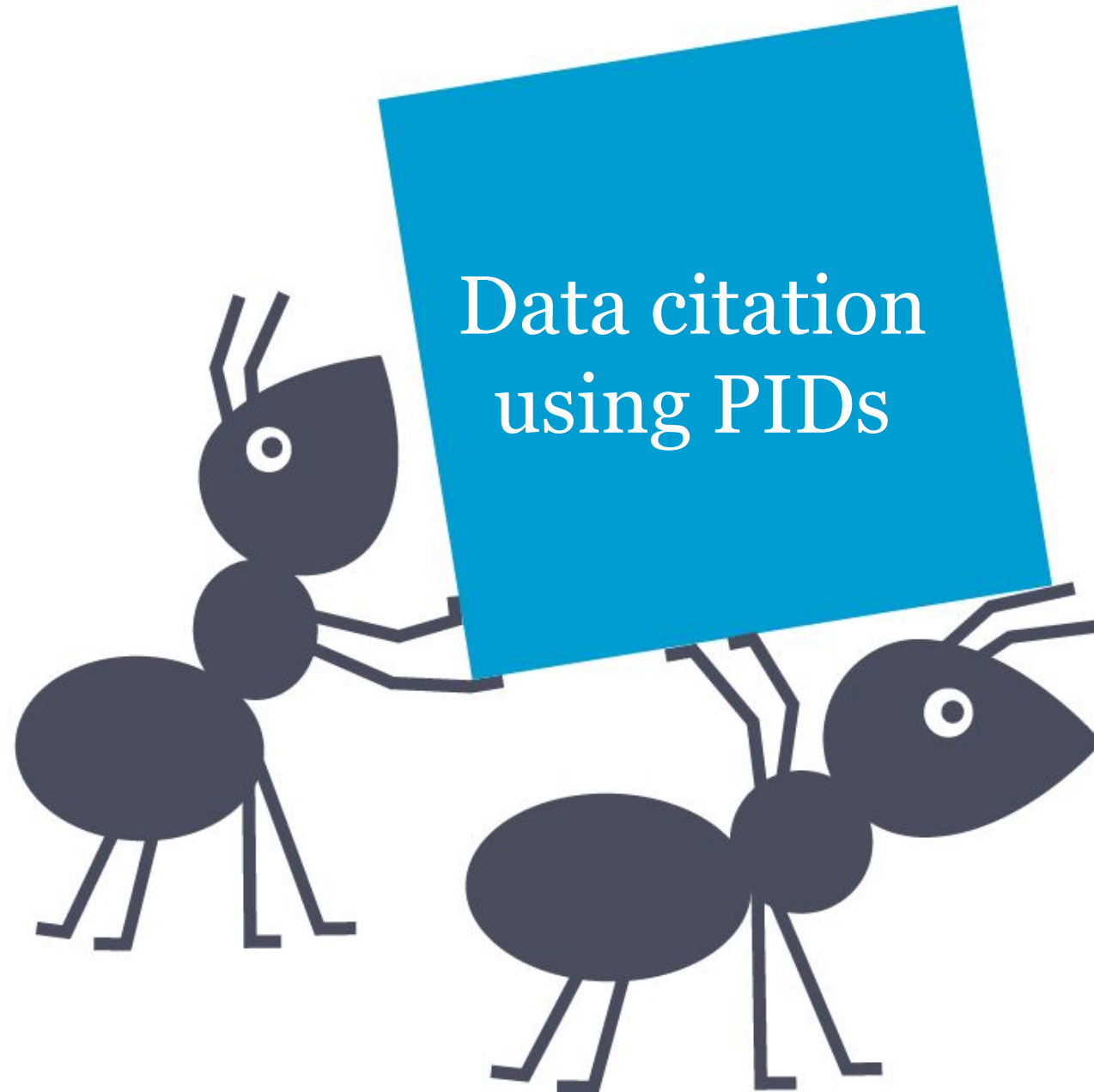
Transcripts

great
awesome *you're amazing* *thumbs up*
brilliant *excellent* *wonderful* **great**
bravo *very good* **Wow** *congratulations*
nice work *perfect* **Nice Work** *super* *awesome*
congratulations *Wow* **good work** **brilliant**
perfect **wonderful** *perfect* **superb** *nice work*
great **thumbs up** *Wow* **Keep it up!** **excellent**
very good *you're amazing*

The Research data granularity levels examples

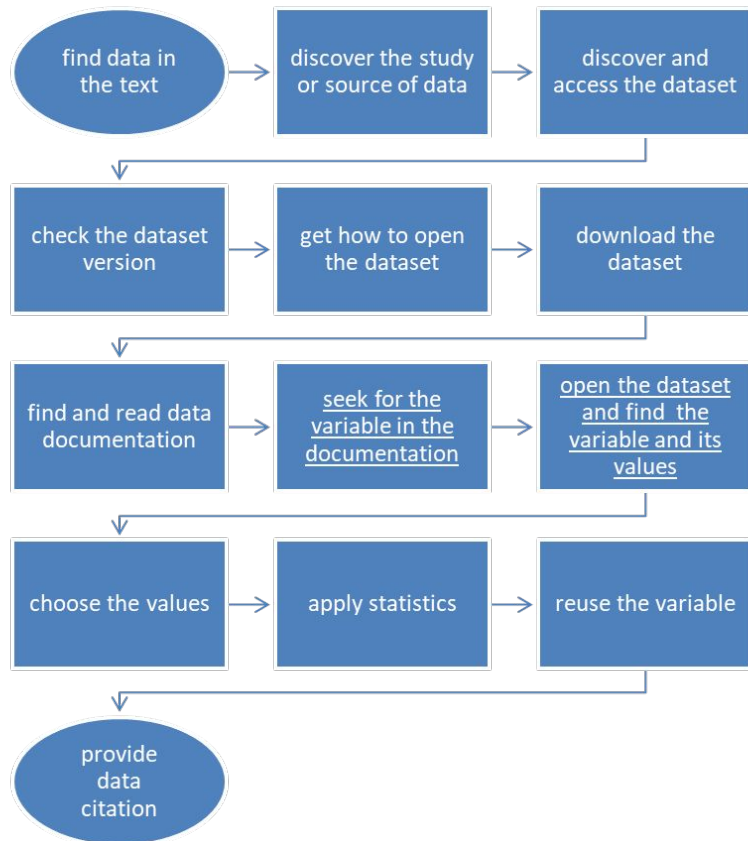
Variables





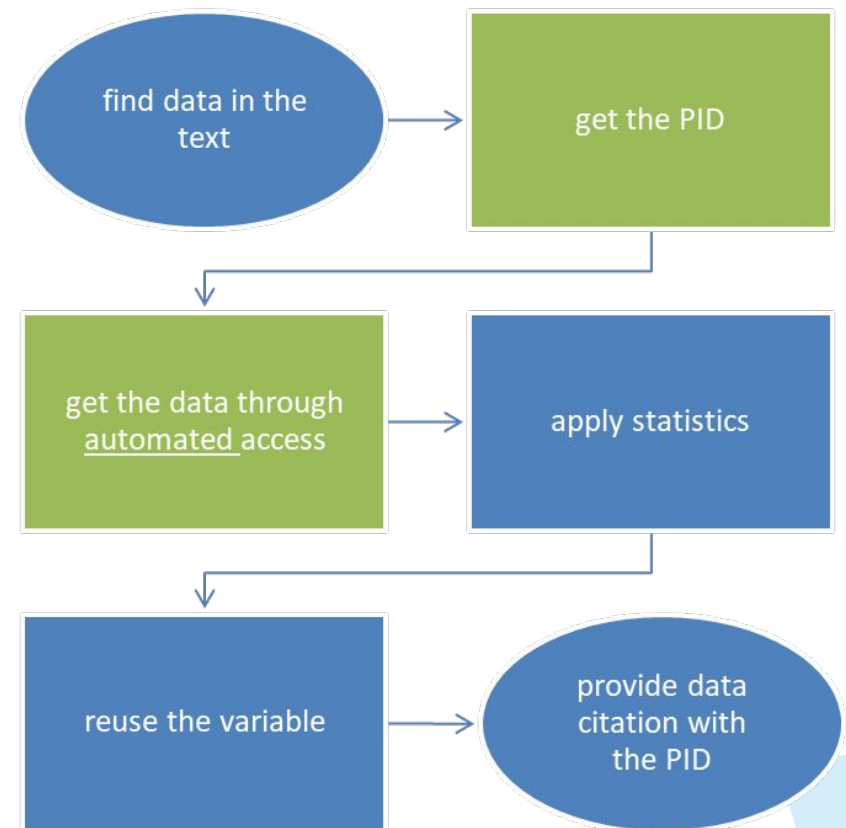
- Finding and getting the variable data using or not a PID:

Process of accessing a variable without PID



X

Process of accessing a variable with a PID



Assigning PIDs for institutions such as:



Research Data Centres (RDCs) potential users





The service
FAIR maturity
level assessment

The service FAIR maturity level assessment: Criteria

- We assessed the service under the FAIR Data Maturity Model (RDA Working Group on FAIR Data Maturity Model, 2020, see doi: [10.15497/rda00050](https://doi.org/10.15497/rda00050))

FAIR Data Maturity Model
Specification and Guidelines



DOI: [10.15497/rda00050](https://doi.org/10.15497/rda00050)

Co-chairs: Edit Herczog, Keith Russell, Shelley Stall

Published: 25th June 2020

Abstract: Findability, Accessibility, Interoperability and Reusability – the FAIR principles – intend to define a minimal set of related but independent and separable guiding principles and practices that enable both machines and humans to find, access, interoperate and re-use data and metadata. The FAIR principles were defined in 2016 in an article by Mark Wilkinson et. al1. FORCE112 and GO FAIR3 provide further information on the principles. The principles have to be considered as

The service FAIR maturity level assessment: Criteria

- The framework consists of **3 indicators classes**: Essential, Important, and Useful
- The sum of them is organized into **five levels**, according to the present indicator in each category
- When distributing the indicators per FAIR area, the principle of **Accessibility** and **interoperability** holds the majority of Essential and Important criteria for FAIRness

3 indicators classes in five levels

FAIR Data Maturity Model: evaluation framework		Level 1	Level 2	Level 3	Level 4	Level 5
Essential	20	20	20	20	20	20
Important	14		7	14	14	14
Useful	7				3	7
Total	41	20	27	34	37	41

Indicators according to the FAIR Principles

Distribution of priorities per FAIR area					
Principle	Findable	Accessible	Interoperable	Reusable	Total
Essential	7	8	0	5	20
Important	0	3	7	4	14
Useful	0	1	5	1	7
Grand Total	7	12	12	10	41

The service FAIR maturity level assessment: Methodology

- Applied the **stricter** evaluation method on each indicator, assessing them by passing or failing **binary answers**
- This approach was selected because the PID registration service is a widening solution to an established service through da|ra (da-ra.de)
- Link to assessment data:



Measure 5.1: PID Service for variables	Present	Not present		
FAIR Data Maturity Model: criteria framework	Pass	Fail	Evidence	Comments
RDA-F1-01M Metadata is identified by a persistent identifier	Pass		It has a variable PID assigned	Metadata and data is identified via DOI
RDA-F1-01D Data is identified by a persistent identifier	Pass		It has a variable PID assigned	Metadata and data is identified via DOI
RDA-F1-02M Metadata is identified by a globally unique identifier	Pass		Handle standard provides globally unique identifier	Metadata and data is identified via DOI
RDA-F1-02D Data is identified by a globally unique identifier	Pass		Handle standard provides globally unique identifier	Metadata and data is identified via DOI
RDA-F2-01M Rich metadata is provided to allow discovery	Pass		A metadata scheme is present to comply with the minimum metadata	Metadata is documented in DDI Lifecycle 3.2
RDA-F3-01M Metadata includes the identifier for the data	Pass		It includes the DOI of the study in which the variable to register appears	The DOI is part of the metadata
RDA-F4-01M Metadata is offered in such a way that it can be harvested and indexed	Pass		The metadata is in fact harvested and indexed at Gesis Search and/or other institutional repository as a service user	Metadata can be harvested via OAI-PMH

The service FAIR maturity level assessment: Results

- The PID registration service **passed 33** indicators and failed 8
- The results for each level were in the range from **80% to 100%**

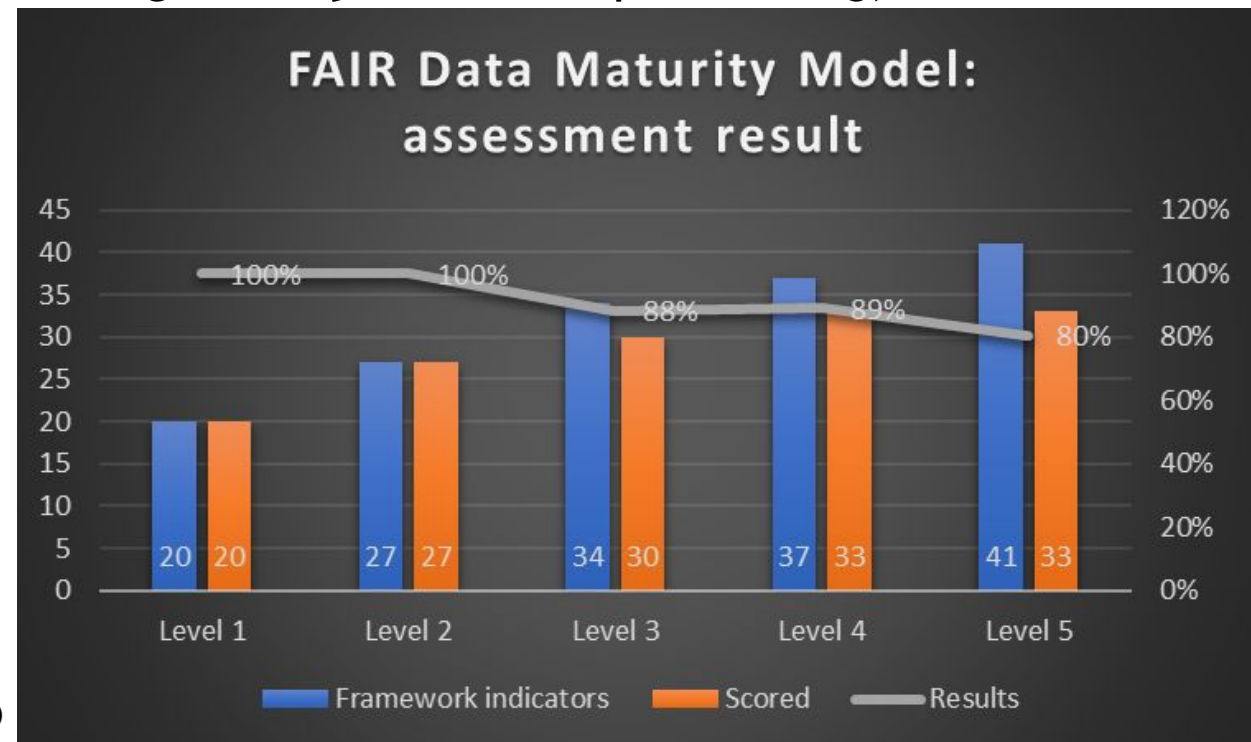


FAIR Data Maturity Model: assessment		
	Pass	Fail
Essential	20	0
Important	10	4
Useful	3	4
Total	33	8

Framework	Level 1	Level 2	Level 3	Level 4	Level 5
Essential	20 / 20	20 / 20	20 / 20	20 / 20	20 / 20
Important		7 / 7	10 / 14	10 / 14	10 / 14
Useful				3 / 3	3 / 7
Achieved indicators	20/20	27 / 27	30 / 34	33 / 37	33 / 41
Scored	20	27	30	33	33
Results	100%	100%	88%	89%	80%

The service FAIR maturity level assessment: Results

- The results demonstrate **outstanding achievements at levels 1 and 2**, marking **100%** on the assessment measure
- The service achieves **88%** compliance at level 3 and **89%** at level 4. At level 5, the results show **80%** of passed indicators
- The service meets **all** indicators classified as **essential**
- The failed indicators concerned with **automatic features**, including references and/or qualified references to other data, and data is accessed automatically (i.e., by a computer program)



The service FAIR maturity level assessment: outcomes

FAIR maturity level assessment of the PID service **confirm the initial assumption that:**

- PIDs on variable level improve/simplify FAIR data management because it:
 - Enables safe data **citation**;
 - Improves **findability**;
 - Fosters **reuse**;
 - Favors **reproducibility**;
- The **failed** indicators so far (**automatic features**) are **feasible to be implemented** in the future since it requires only the PID assigned to the variable and a code/do-file (i.e., by a computer program) designed to **get the data automatically**. It is a real potential advantage for the data provider and data users.

The service FAIR maturity level assessment: data source

The list of indicators, their assessment with related evidence, and comments are available at the link below.

<https://docs.google.com/spreadsheets/d/1R9aoimBwoVdP5yxyA3h7mguGB6vIBaU7/edit?usp=sharing&oid=105103210002302942928&rtpof=true&sd=true>

Janete Saldanha Bach, Claus-Peter Klas and Peter Mutschke. 2022. Application of ‘RDA FAIR Data Maturity Model’ to assess the PID registration service in terms of FAIRness. In *CESSDA, Webinar on User Experience with FAIR Evaluation Tools and Services*. Cologne, Germany, 11.10.2022, 28 slides.



Thank you

The service is part of KonsortSWD project deliverable, NFDI funding number 442494171

