



FAIRsFAIR
Fostering Fair Data Practices in Europe

F-UJI FAIR Evaluator

Capabilities and Limitations of FAIR assessments

Robert Huber rhuber@uni-bremen.de

#FAIRification step 5



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>

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"

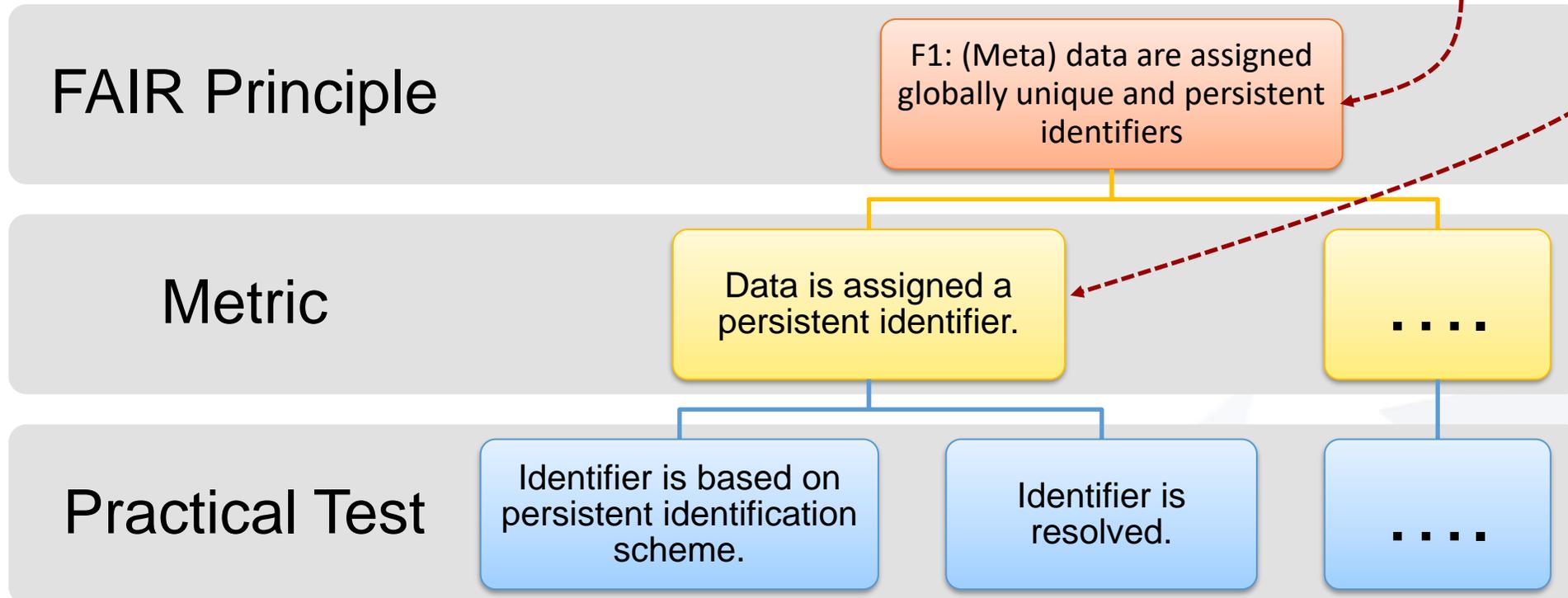
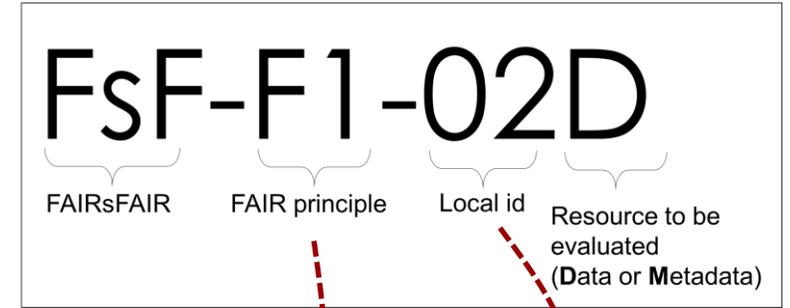
Object Assessment Metrics v0.4

We would love to hear your feedback!

<https://fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments>



From Principles to Practical Tests



Huber, Robert, Cepinskas, 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>



Defining tests – based on what?

Principle	Metrics	Practical Tests	Rationale
F1	FsF-F1-02D Data is assigned a persistent identifier	<ul style="list-style-type: none"> A data identifier is specified based on a commonly accepted persistent identifier scheme suitable for research data. The identifier is web-accessible, i.e., it resolves to a landing page with metadata of the data object. 	<ul style="list-style-type: none"> EOSC PID policy: globally unique, persistent, resolvable, managed (Valle et al. 2020) DataCite identifier type vocabulary (DataCite Metadata Working Group 2019) + identifiers.org (No authoritative registry of valid persistent identifiers exists)
F2	FsF-F2-01M Metadata includes descriptive core elements to support data findability.	<ul style="list-style-type: none"> Some metadata (at all) has been made available via common (web) standards. Minimum core citation metadata is specified (creator, title, publication date, publisher, and identifier) Minimum core descriptive metadata is specified (creator, title, publisher, publication date, summary, keywords, identifier) through appropriate metadata fields. 	<ul style="list-style-type: none"> OAIS reference model (ISO 14721:2012): ‘Findable’ => OAIS descriptive metadata Data citation: <ul style="list-style-type: none"> <i>Recommendations</i> of Force11, ESIP, IASSIST, DataCite: Data description: <ul style="list-style-type: none"> <i>Standards/Recommendations:</i> EOSC Datasets Minimum Information, DataCite Metadata Schema, W3C Recommendation Data on the Web, Data Catalog Vocabulary (DCAT-2) <i>Communality analysis</i> of common domain agnostic metadata standards

Huber, Robert, Cepinskas, 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>

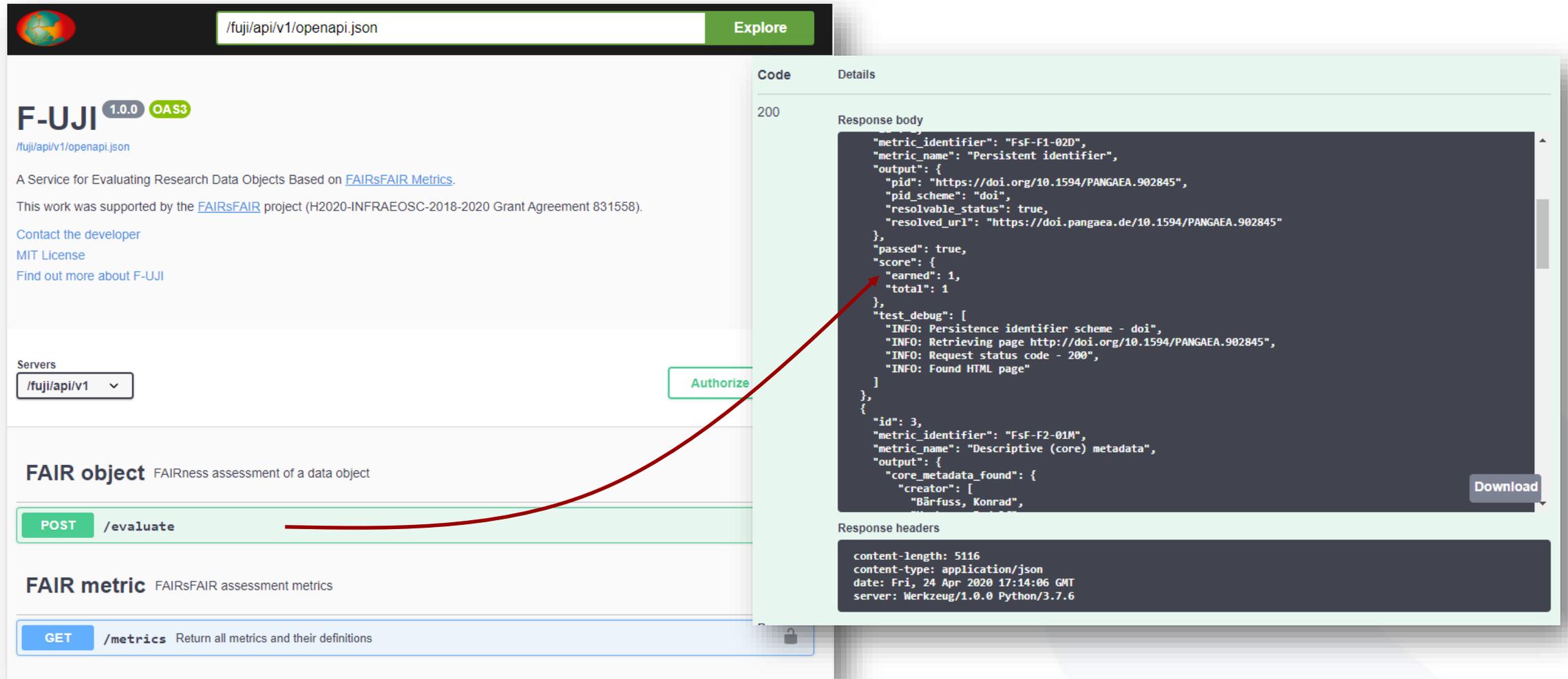


<https://github.com/pangaea-data-publisher/fuji>

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



F-UJI – An Automated FAIR Data Assessment Tool



The screenshot displays the F-UJI API interface. At the top, there is a search bar with the URL `/fuji/api/v1/openapi.json` and an **Explore** button. Below this, the API is identified as **F-UJI 1.0.0 OAS3**. The main content area shows the API description, including a link to the `/fuji/api/v1/openapi.json` file, a description as a service for evaluating research data objects, and support information from the FAIRSFAR project. A **Servers** dropdown menu is set to `/fuji/api/v1`. Below the description, there are sections for **FAIR object** and **FAIR metric**. The **FAIR object** section shows a **POST /evaluate** endpoint. The **FAIR metric** section shows a **GET /metrics** endpoint. A **Authorize** button is visible. On the right, a **Code** panel shows the response body for a `200` status code. The response is a JSON object with the following structure:

```
Response body
{
  "metric_identifier": "FsF-F1-02D",
  "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 http://doi.org/10.1594/PANGAEA.902845",
    "INFO: Request status code - 200",
    "INFO: Found HTML page"
  ]
},
{
  "id": 3,
  "metric_identifier": "FsF-F2-01M",
  "metric_name": "Descriptive (core) metadata",
  "output": {
    "core_metadata_found": {
      "creator": [
        "Bärfuss, Konrad",
        "Bärfuss, Konrad"
      ]
    }
  }
}

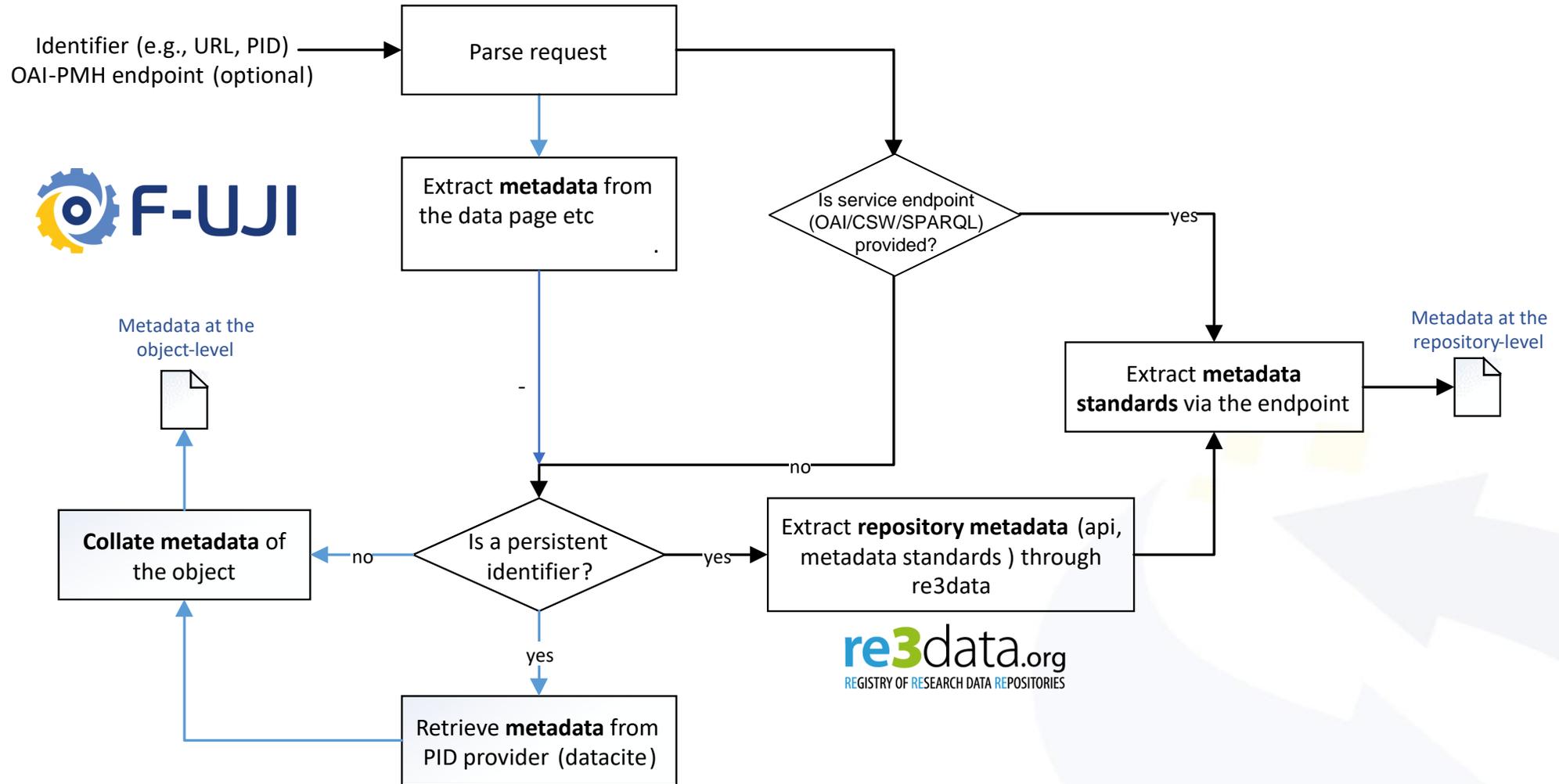
```

The **Response headers** section shows:

```
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
```

A red arrow points from the **Authorize** button to the **Response body** section. A **Download** button is located at the bottom right of the response body area.

High Level Flow (Data Gathering)



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).

Save JSON New

FAIR level: initial

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

DataCite support: enabled

Metric Version: metrics_v0.4

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

Software version: v1.3.8

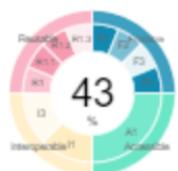
Download assessment results: [\(JSON\)](#)

Save and share assessment results:

Saved assessments:

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

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

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

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": "http://doi.org/10.17863/CAM.14473",
  "pid_scheme": "doi",
  "resolvable_status": true,
  "resolved_url": "https://www.repository.cam.ac.uk/handle/1810/268269"
}
```

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', 'ensembl', 'genome', 'gnd', 'handle', 'lsid', 'pmid', 'pmcid', 'purl', '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/rdf+xml;q=0.5
INFO	Content negotiation accept=text/html, application/xhtml+xml, application/xml;q=0.5, text/xml;q=0.5, application/rdf+xml;q=0.5, status=200
INFO	Found HTML page!
INFO	Object identifier active (status code = 200)
SUCCESS	Persistence identifier scheme -: doi

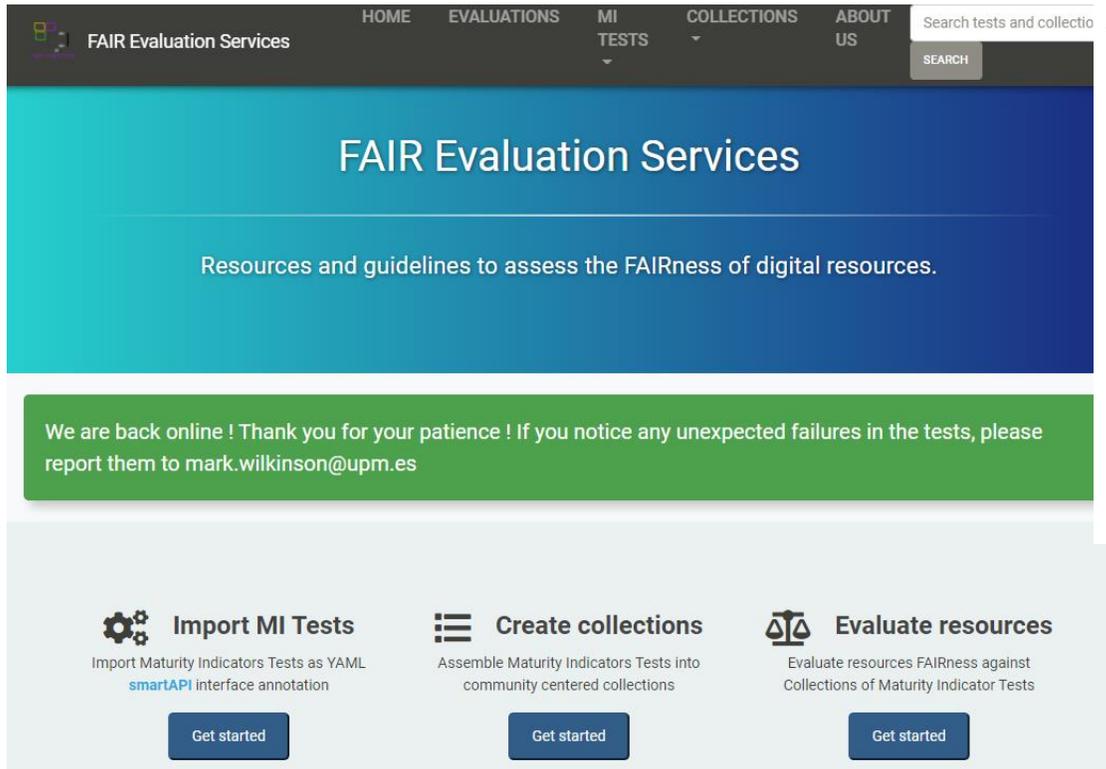
FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability. ✓

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. ✓



Landscape of tools



FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.

We are back online ! Thank you for your patience ! If you notice any unexpected failures in the tests, please report them to mark.wilkinson@upm.es

- Import MI Tests**
Import Maturity Indicators Tests as YAML smartAPI interface annotation
[Get started](#)
- Create collections**
Assemble Maturity Indicators Tests into community centered collections
[Get started](#)
- Evaluate resources**
Evaluate resources FAIRness against Collections of Maturity Indicator Tests
[Get started](#)

<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#/!/#%2F!>

For a comparison, see: Chang Sun, Vincent Emonet, & Michel Dumontier. (2021). Comparison results of FAIR Evaluation tools (V0.1) [Data set]. Semantic Web Applications and Tools for Healthcare and Life Sciences (SWAT4HCLS), Leiden, Netherlands. Zenodo.

<https://doi.org/10.5281/zenodo.5539823> - Paper to come

AUTOMATED EVALUATIONS

FAIRMetrics APIs

Metadata quality checks

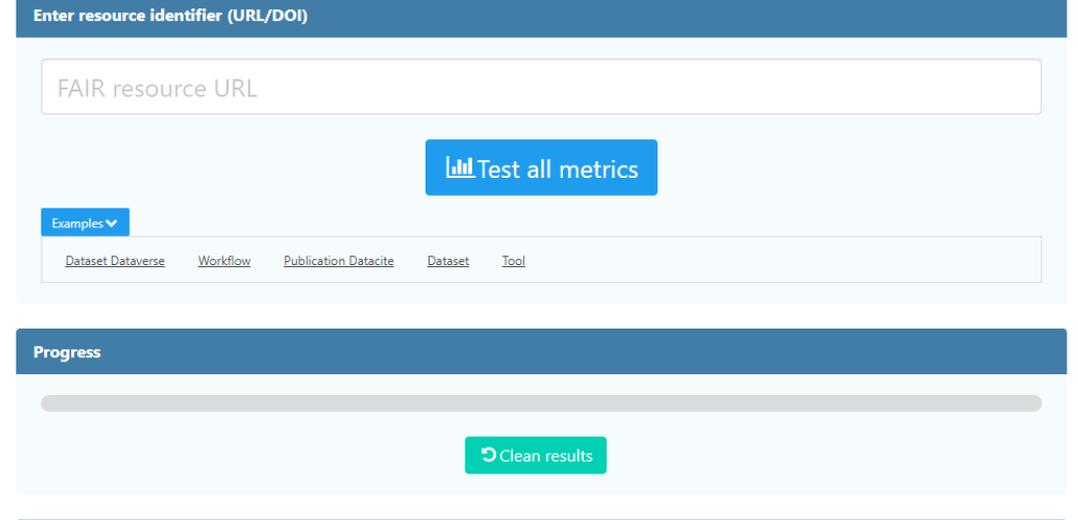
ABOUT

About & Feedbacks

Usage statistics

Disclaimer: FAIR-Checker is under active development and is not intended to be used yet in production environment.

How FAIR is my resource ?



Enter resource identifier (URL/DOI)

FAIR resource URL

[Test all metrics](#)

Examples ▾

[Dataset](#) [Dataverse](#) [Workflow](#) [Publication](#) [Datacite](#) [Dataset](#) [Tool](#)

Progress

[Clean results](#)

https://fair-checker.france-bioinformatique.fr/base_metrics

Limitations – Comparing results

How to compare assessment results?:

- Different metrics
- Different tests
- Different software
- Different results
- Different FAIR estimates

	A	B	C	D	E	F	G	H
1			[10.1594/PANGAEA.9		[https://www.kaggle.		[https://data.rivm.nl/	
2			Fuji	Evaluator	Fuji	Evaluator	Fuji	Evaluator
3	Findable							
4	-	Gen2_MI_F1A	-	✓	-	✓	-	✓
5	-	Gen2_MI_F1B	-	X	-	X	-	X
6	FsF-F1-01D	Gen2_MI_F1B	✓	X	✓	X	✓	X
7	FsF-F1-02D		✓		X		X	
8	FsF-F2-01M	Gen2_MI_F2A	✓	✓	✓	✓	✓	✓
9	-	Gen2_MI_F2B	-	✓	-	✓	-	✓
10	FsF-F3-01M	Gen2_MI_F3	✓	✓	✓	✓	✓	✓
11	-	Gen2_MI_F3	-	X	-	X	-	X
12	FsF-F4-01M	Gen2_MI_F4	✓	✓	✓	✓	X	X

Chang, Vincent, & Michel. (2021). Comparison results of FAIR Evaluation tools (V0.1) [Data set]. Semantic Web Applications and Tools for Healthcare and Life Sciences (SWAT4HCLS), Leiden, Netherlands. Zenodo. <https://doi.org/10.5281/zenodo.5539823>

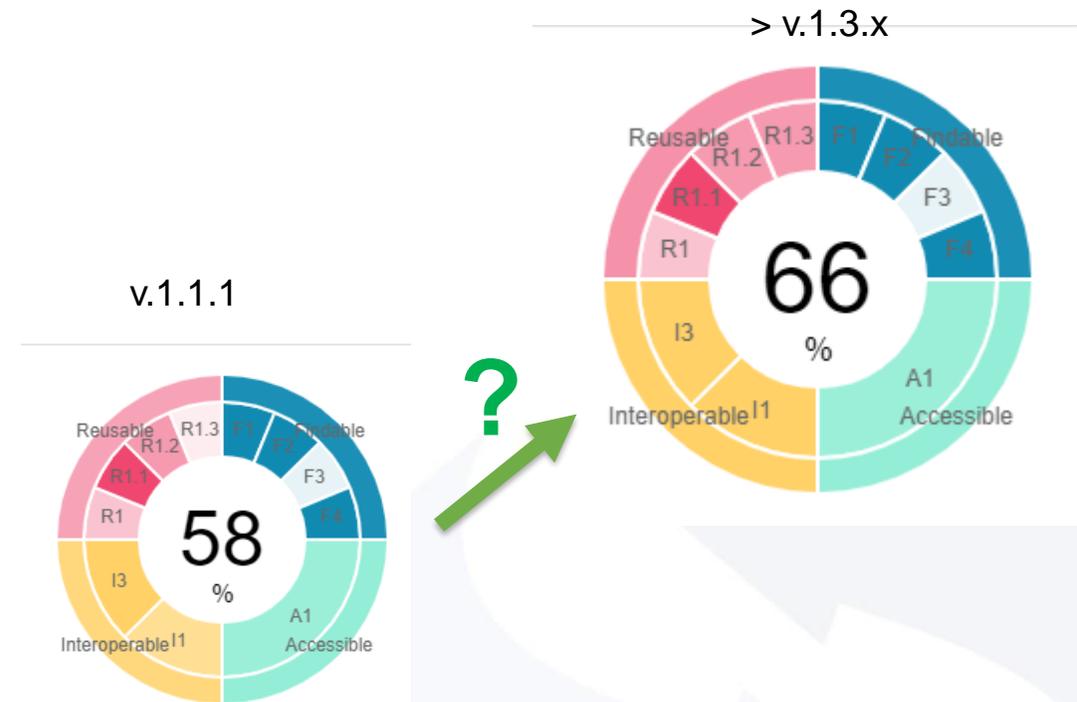
- **Lack of standard samples**
- **Lack of calibration procedures**



Limitations – Interpreting results

How to interpret or benchmark assessment results?:

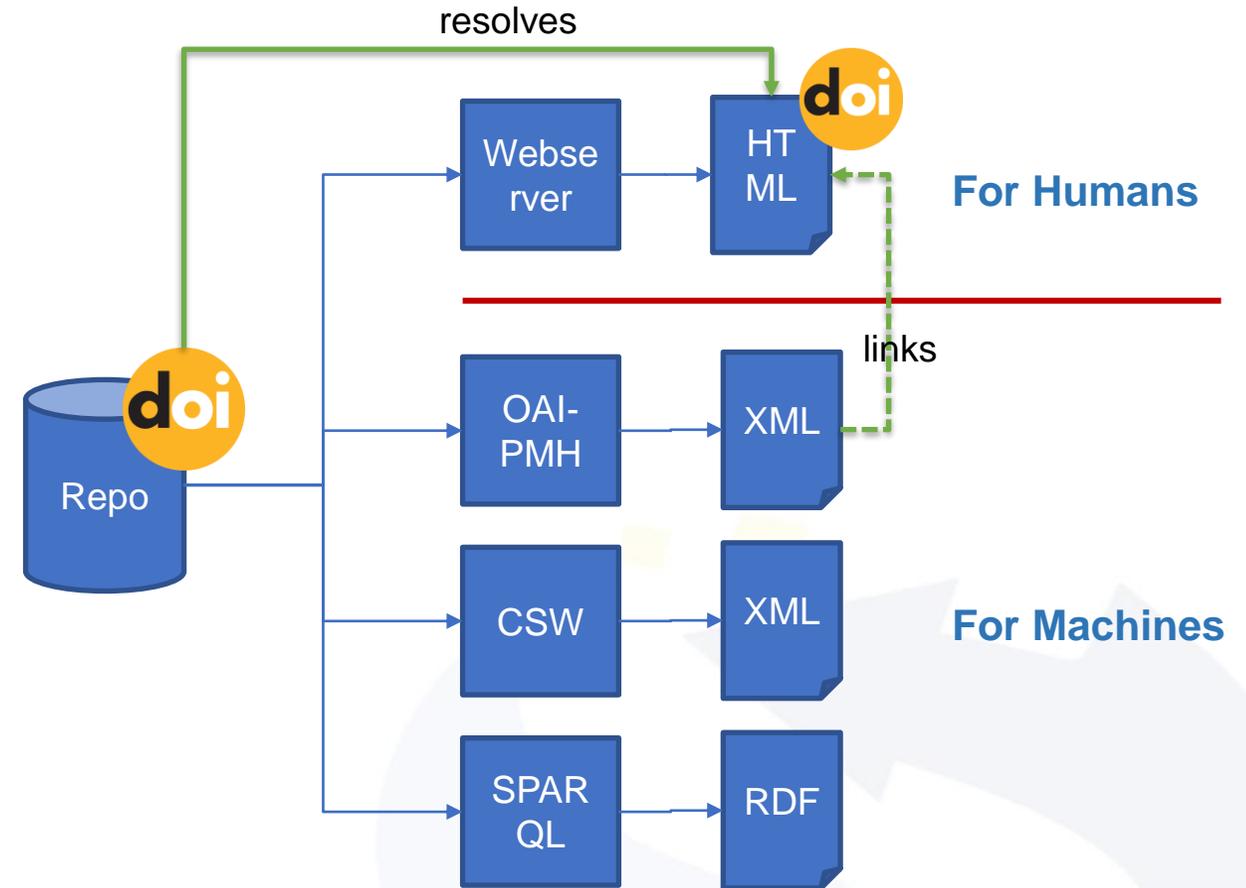
- Sample may change
 - PID redirection change,
 - Repo software change
 - FAIR improvement
 - etc..
- Procedure may change
 - Metrics change
 - Scoring change
 - Software versions
 - etc..
- **Lack of standard samples**
- **Lack of calibration procedures**



<http://doi.org/10.22033/ESGF/CMIP6.4397>

Limitations – Lack of Connectivity

- FAIR architectures rely on standard (interfaces)
- Many PIDs point to landing pages only
- PIDs not supported by standards
 - OAI-PMH
- PIDs and Interfaces often disconnected: cannot be discovered



Limitations – Lack of stability

- F-UJI uses third party resources:
 - Datacite
 - Re3data
 - Mendeley data
- **These may fail...**

High load amongst content negotiation service.

Incident Report for DataCite

Resolved

We are still receiving a high load of requests but we are now fulfilling due to increased capacity, so this issue is being marked resolved.

Requests for DataCite DOIs should be unaffected going forward, requests for Crossref DOIs against DataCite Content negotiation directly may occasionally return 404s when it is unable to process.

It is preferred if possible to use Content Negotiation via doi.org and you will be redirected as appropriate to the registration agencies content negotiation service as appropriate.

Posted 14 days ago Jan 24, 2022 - 09:20 UTC

PIXAR



AWWW...DON'T CRY.

It's just a 404 Error!

What you're looking for may have been misplaced in Long Term Memory.



Limitations – Handling disciplinary diversity

- Example: Geographic Information – Metadata
- Very complex!
- Several versions for XML:
e.g. 19139 , 19115-3
- Community profiles
 - INSPIRE
- Many different representations of the same standard
- Hard to parse

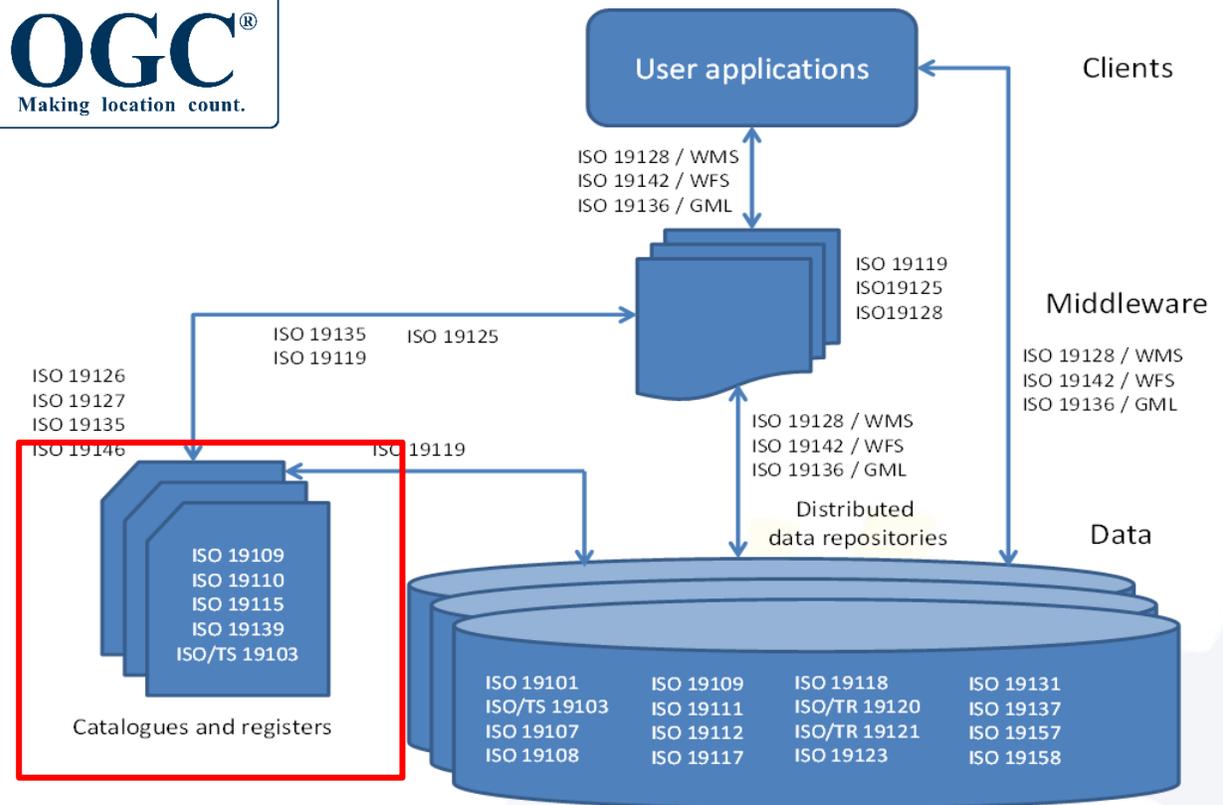


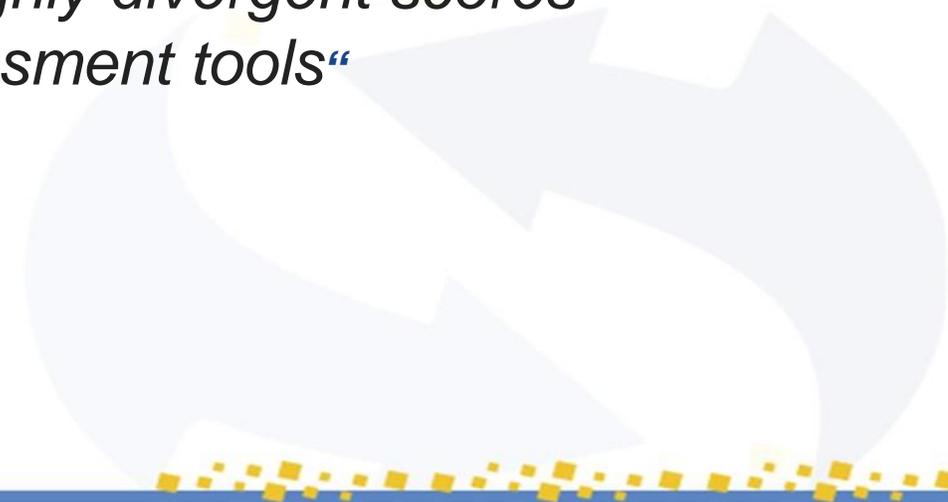
Image source: https://eo4geo.sbg.ac.at/KULeuven/Technical_Introduction_SDI/Images_Module01/Standards_for_SDI.png

Limitations – How to deal?

FAIR Evaluation stakeholder meeting (Feb. 7 & 10, 2022)



*„Key FAIR Evaluation stakeholders together to discuss **workflows for metadata gathering**. This is prompted by the highly divergent scores currently being issued by the various FAIR assessment tools“*



Thank You

Task 4.5:

Anusuriya Devaraju, Robert Huber,
Mustapha Mokrane, Jerry de Vries,
Patricia Herterich, Linas Cepinkas, Vesa
Akerman, Joy Davidson, Herve L'Hours.

FAIR-Aware

Let's assume you have research data almost ready for uploading to a repository: do you already know how you and the repository can work together to make the data as findable, accessible, interoperable and reusable (FAIR) as possible? By guiding you through the assessment process, the FAIR-Aware tool can help you to better understand the FAIR Principles and how making data FAIR can increase the potential value and impact of your data.

FAIR-Aware is an online tool developed by the FAIRSFAR project. The tool is not meant to give you a score for the FAIRness of a specific dataset. You should, however, have a target dataset in mind to be able to answer the questions and complete the assessment.

The assessment starts with a few questions 'about you' followed by 10 questions about FAIR. After you answer each question additional information and guidance will be displayed. The majority of the questions will help you assess your current level of awareness about what actions are needed to make data FAIR. At the end, Your feedback will help us improve FAIR-Aware and make it as user-friendly as possible. You will need between 10 and 30 minutes to complete the assessment depending on your familiarity with the subject and issues covered.

The FAIRSFAR Team (DANS, DCC, UniHB)

Find out more about FAIRSFAR on the project's website. If you have any questions, drop us an e-mail.

About you

Which research domain do you work in? Domain ▾

Which of the following describes your role? Please select all that apply.

- | | |
|--|------------------------------------|
| <input type="checkbox"/> Researcher | <input type="checkbox"/> Funder |
| <input type="checkbox"/> Policy maker | <input type="checkbox"/> Publisher |
| <input type="checkbox"/> Research support (e.g. data steward, curator, data manager, librarian, information technology professional) | <input type="checkbox"/> Other |

Which of the following types of organisations best describe your employer? Please select all that apply.

- | | |
|--|---------------------------------------|
| <input type="checkbox"/> Research Infrastructure/eInfrastructure (e.g. data repository, service provider, library) | <input type="checkbox"/> Funding Body |
| <input type="checkbox"/> University or Research Performing Organisation | <input type="checkbox"/> Publisher |
| <input type="checkbox"/> Research Performing Organisation | <input type="checkbox"/> Industry |
| <input type="checkbox"/> Government | <input type="checkbox"/> Other |
| <input type="checkbox"/> eInfrastructure (e.g. repository or scientific data provider) | |

FAIR questions

FINDABLE

Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository? Yes No

Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others? Yes No

