



**FAIRSFair**  
Fostering Fair Data Practices in Europe

## F-UJI – A FAIR data assessment tool

Robert Huber  
rhuber@uni-bremen.de

ExPaNDS Symposium for Librarians and data managers

# 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

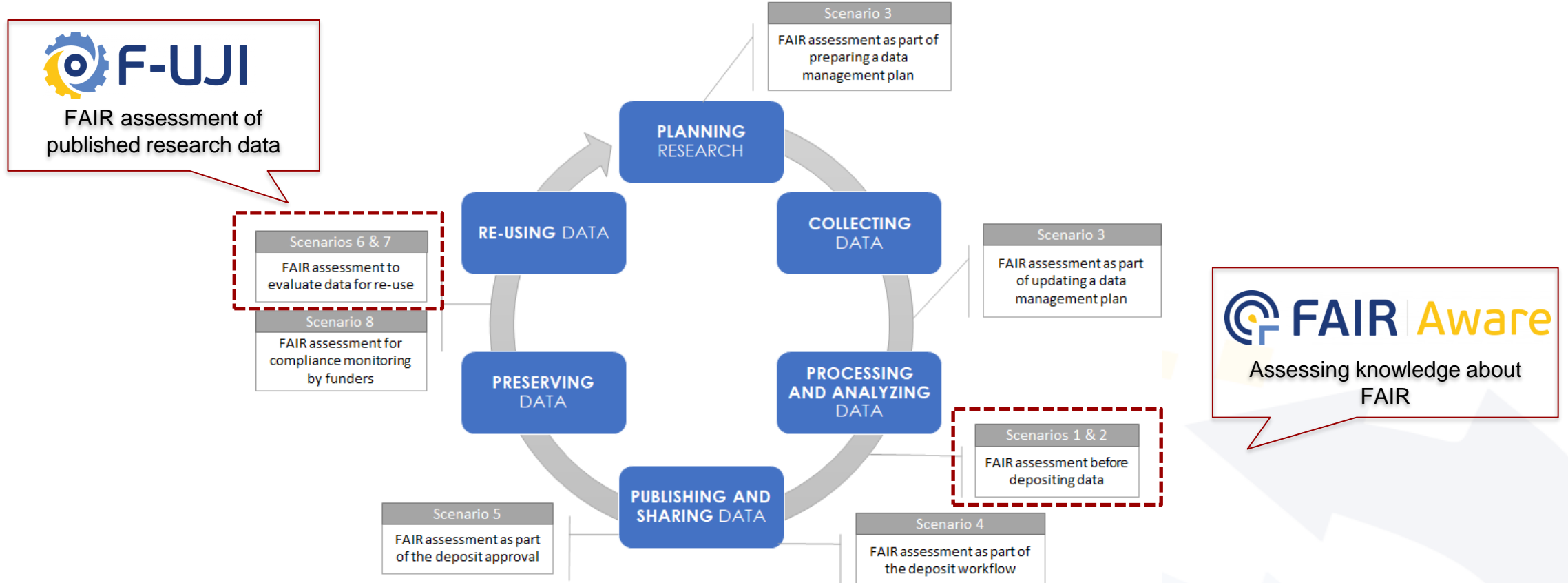
Rec. 25: Implement FAIR metrics to monitor uptake



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

For more information, see D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories, <https://doi.org/10.5281/zenodo.3678715>



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

<sup>1</sup>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

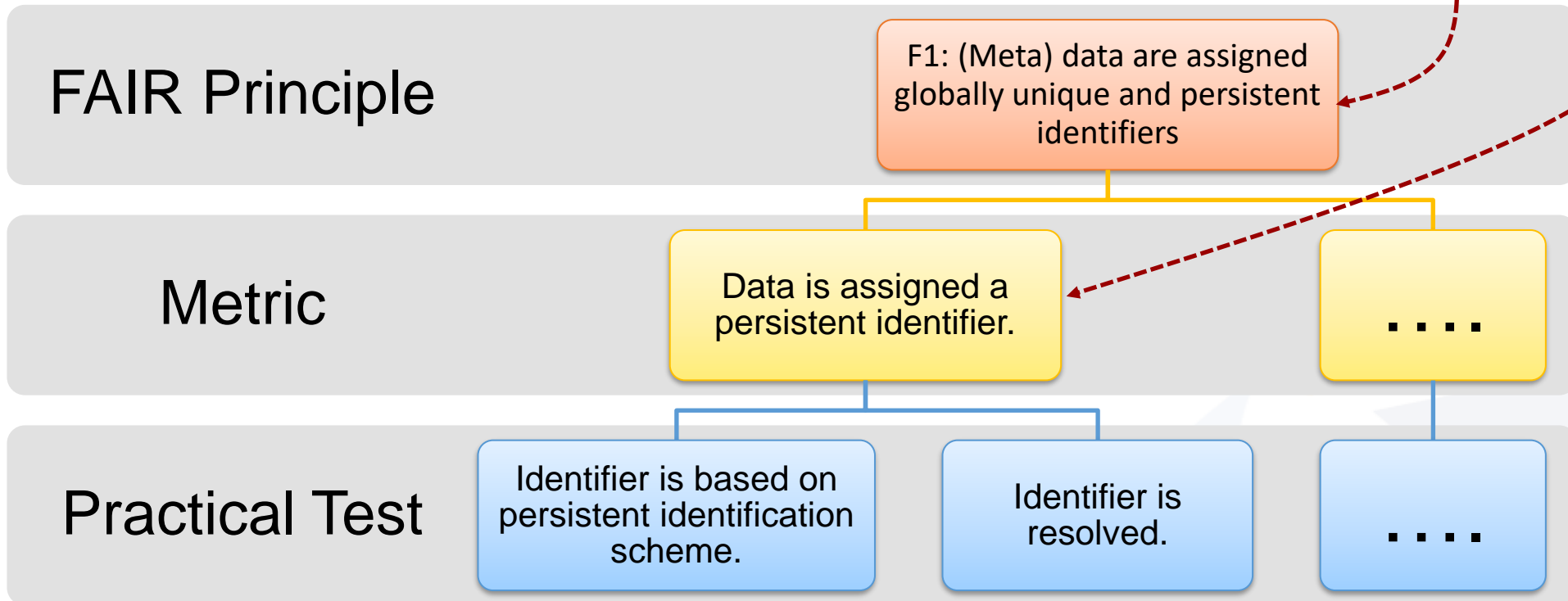
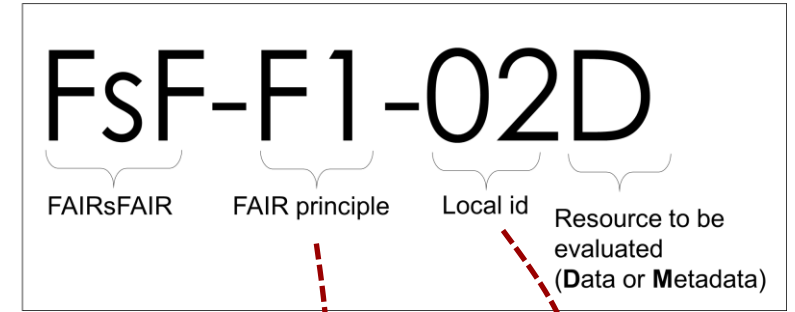
Please [login & comment below](#) citing in the subject line the Metric Identifier No. you are referring to - e.g. "FsF-R1.3-01M"

# 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>



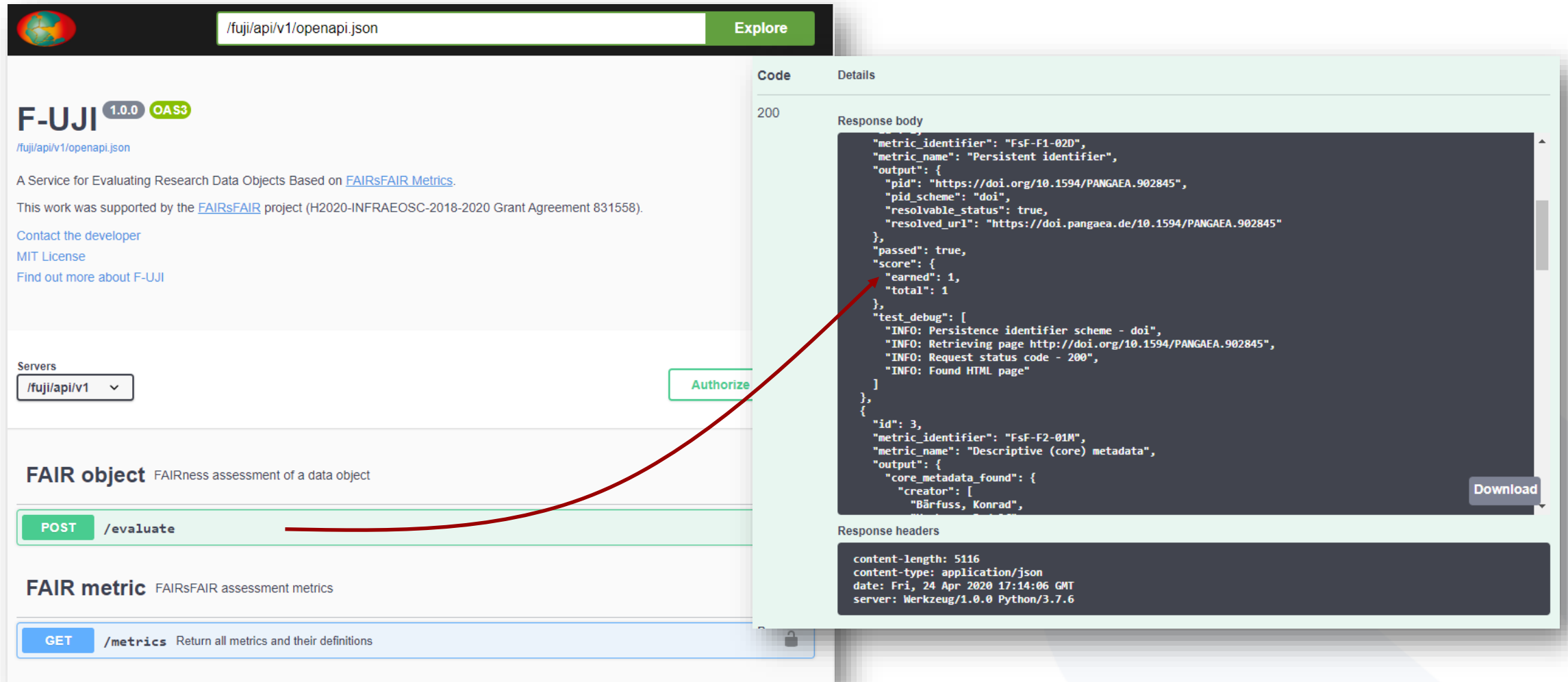


<https://github.com/pangaea-data-publisher/fuji>

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

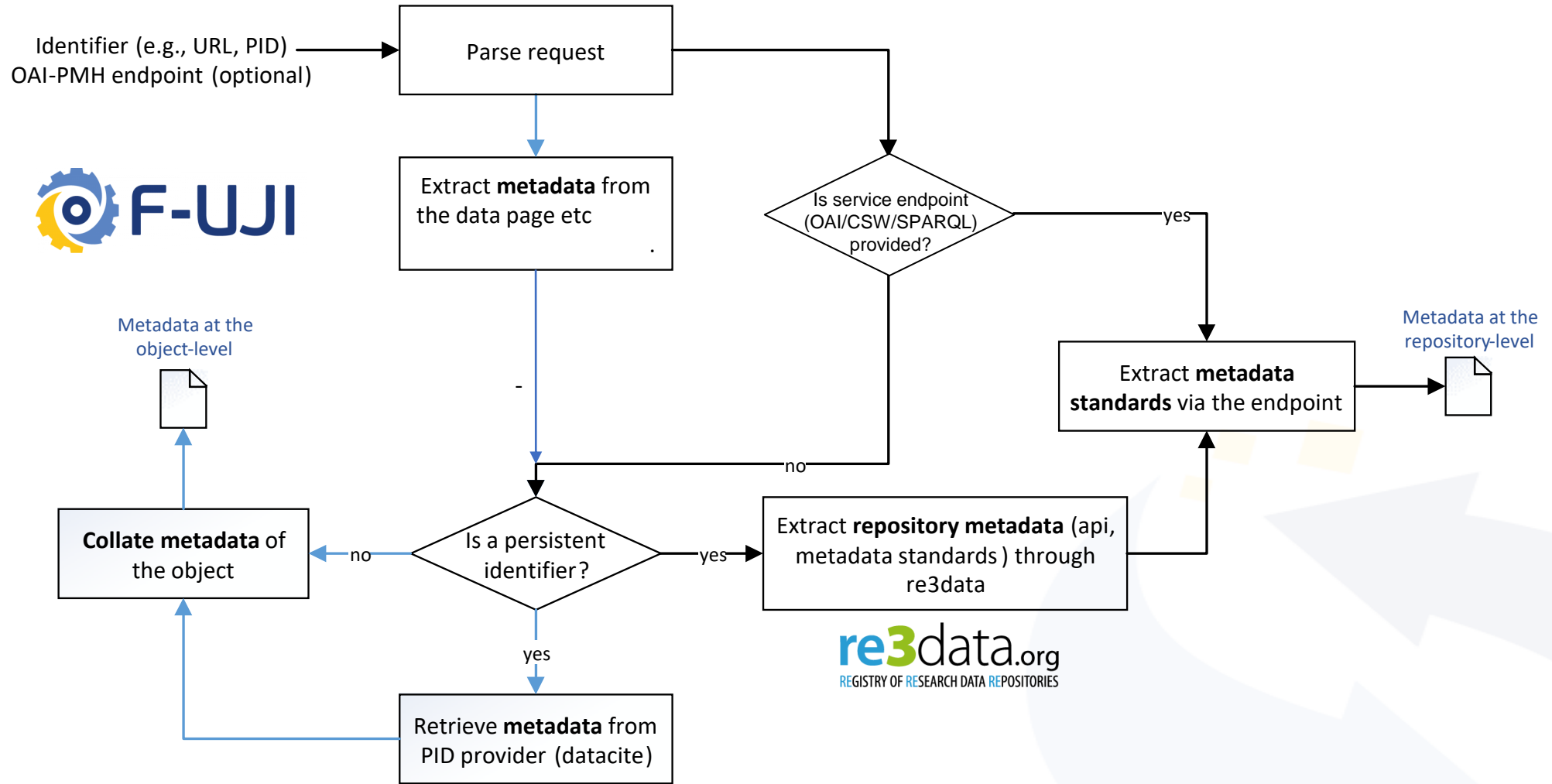


# F-UJI – An Automated FAIR Data Assessment Tool



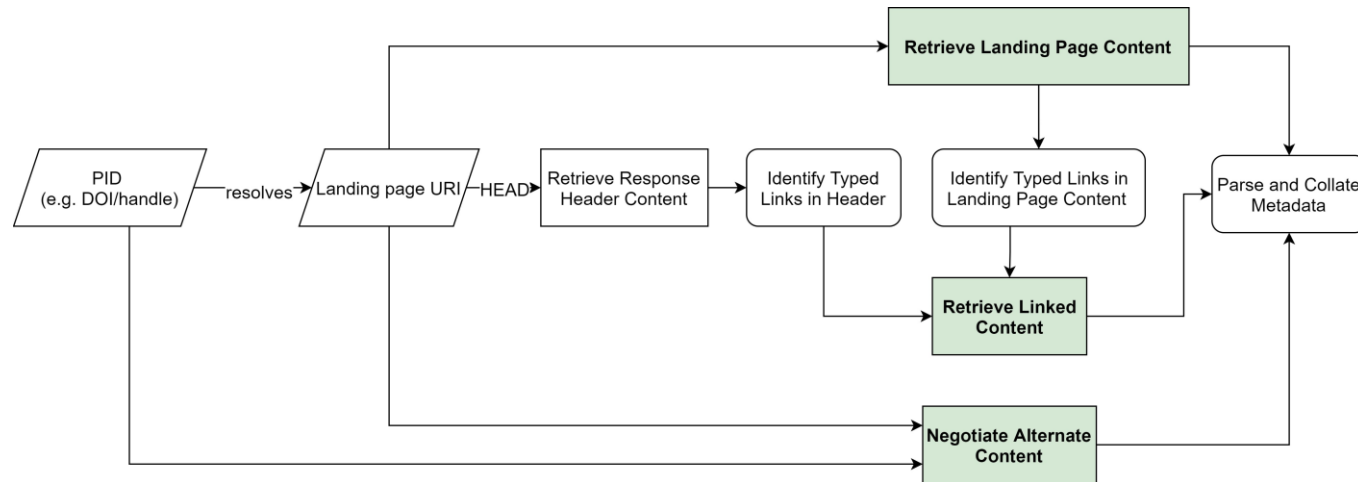
The screenshot displays the F-UJI API interface. At the top, the URL `/fuji/api/v1/openapi.json` is entered in the search bar, with an **Explore** button. The main content area shows the **F-UJI 1.0.0 OAS3** API specification. Below this, there are links for `/fuji/api/v1/openapi.json`, a description of the service, and contact information. A **Servers** dropdown menu is set to `/fuji/api/v1`. The **FAIR object** section shows a **POST /evaluate** endpoint. The **FAIR metric** section shows a **GET /metrics** endpoint. A red arrow points from the `/evaluate` endpoint to a detailed view of the response. This view shows a **200** status code and a **Response body** containing JSON data. The JSON data includes a `metric_identifier` of "FsF-F1-02D", a `metric_name` of "Persistent identifier", and a `passed` status of `true`. The `score` object shows `earned: 1` and `total: 1`. The `test_debug` array contains log messages such as "INFO: Persistence identifier scheme - doi", "INFO: Retrieving page http://doi.org/10.1594/PANGAEA.902845", "INFO: Request status code - 200", and "INFO: Found HTML page". The **Response headers** section shows `content-length: 5116`, `content-type: application/json`, `date: Fri, 24 Apr 2020 17:14:06 GMT`, and `server: Werkzeug/1.0.0 Python/3.7.6`. A **Download** button is visible in the bottom right corner of the response details panel.

# High Level Flow (Data Gathering)





# Metadata collation



- **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).

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	<span>moderate</span>
Accessible:	1.5 of 3	<span>initial</span>
Interoperable:	1 of 4	<span>initial</span>
Reusable:	3 of 10	<span>initial</span>

<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	<span>✓</span>
FsF-F1-02D-2	Persistent identifier is resolvable	0.5	3	<span>✓</span>

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. ✓

# Testing & Consultation – the pilots

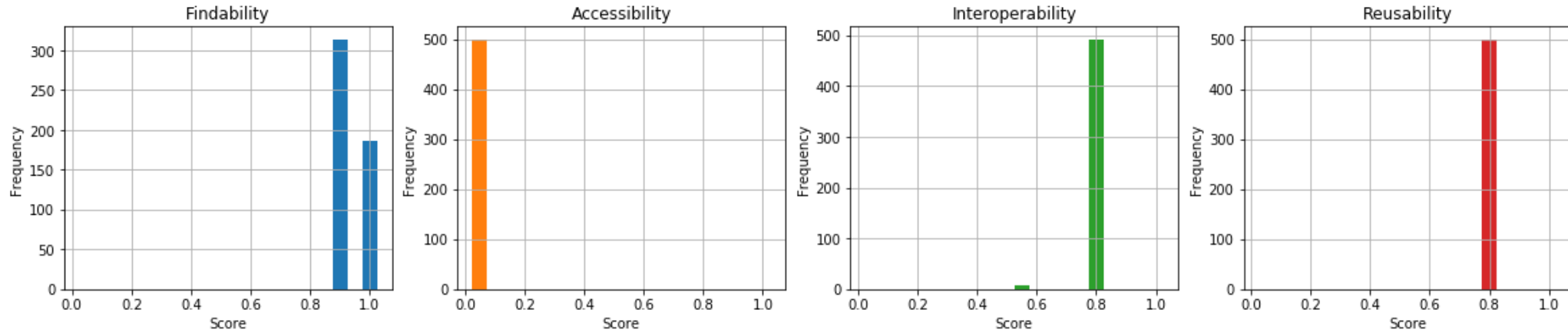
Repository	Certification	Subject Areas	Datasets Evaluated (as of 25.09.2020)	FAIR Data Improvement	Repository Contact
<a href="#">PANGAEA</a>	CoreTrustSeal, WDS Regular Member				
<a href="#">Phaidra-Italy</a>	CoreTrustSeal				
<a href="#">CSIRO Data Portal</a>	CoreTrustSeal				
<a href="#">World Data Centre for Climate (WDCC)</a>	CoreTrustSeal, WDS Regular Member				
<a href="#">DataverseNO</a>	CoreTrustSeal				

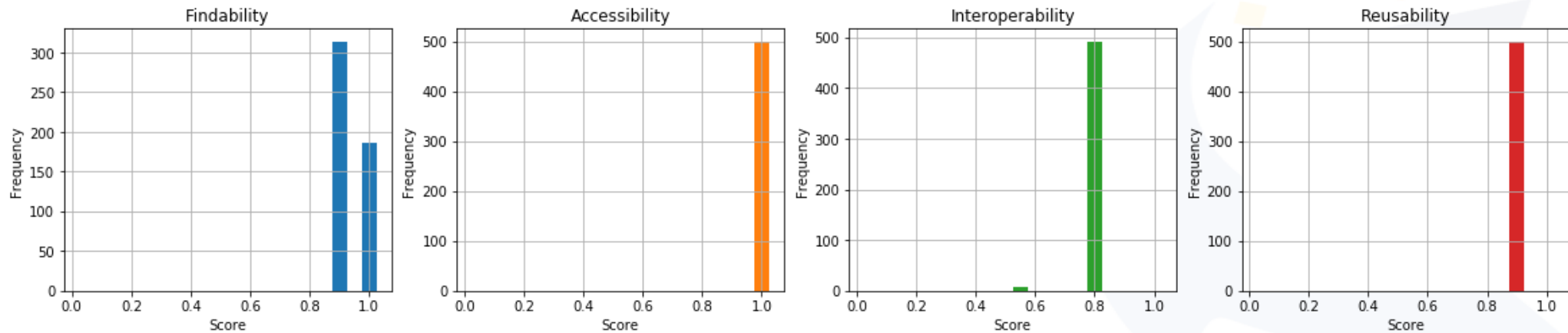
FAIRSFAR Object Metric	Comment	Next Step
<b>FsF-F1-01D</b> Data is assigned a globally unique identifier.	OK	
<b>FsF-F1-02D</b> Data is assigned a persistent identifier.	OK	
<b>FsF-F2-01M</b> Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.	Some datasets do not have a summary/abstract. e.g., <a href="https://doi.org/10.1594/PANGAEA.863175">https://doi.org/10.1594/PANGAEA.863175</a>	Uwe: Most child datasets don't have abstract, but their parents. Abstract is optional field (datacite schema) For new datasets, abstract is recommended especially for datasets with publications. Routine (automatic from sensor) dataset don't have abstract
<b>FsF-F3-01M</b> Metadata includes the identifier of the data it describes.	OK	
<b>FsF-F4-01M</b> Metadata is offered in such a way that it can be retrieved by machines.	OK	
<b>FsF-A1-01M</b> Metadata contains access level and access conditions of the data.	Total score = 1, Earned = 0 NO access information is available in metadata:  <b>Possible ways to include access level metadata are:</b>  Schema.org <ul style="list-style-type: none"> <li>if public use =&gt; <a href="#">isAccessibleForFree</a>    free</li> </ul> (Note: isAccessibleForFree Supersedes free)	Action: Uwe to include access level in the metadata (Done) <input type="checkbox"/>

# Before and After

FAIR Scores of PANGAEA Datasets By Principle (Before Improvement, n=500)



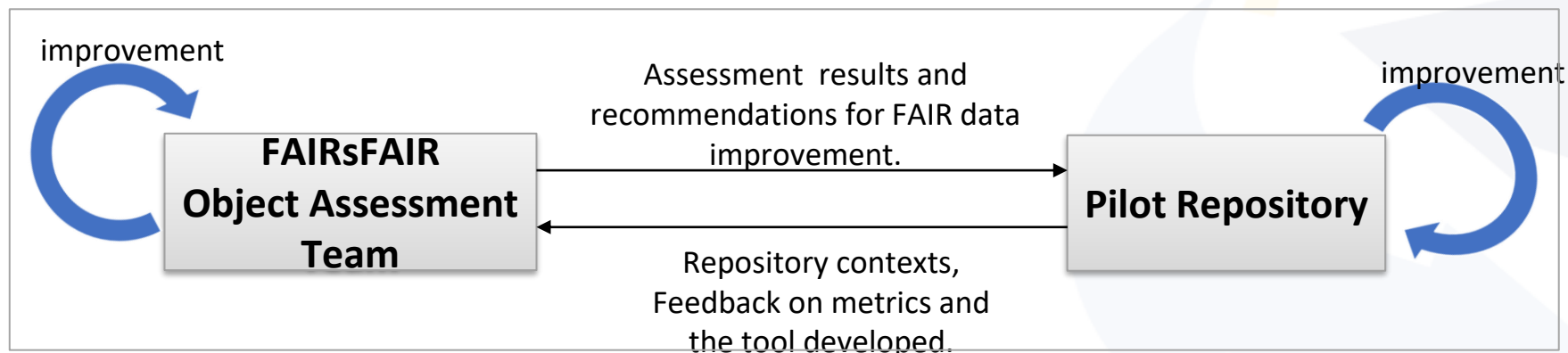
FAIR Scores of PANGAEA Datasets By Principle (After Improvement, n=500)



# Lessons learned

- Automatised FAIR assessment of research data objects is possible
- Supporting a very large and diverse community
  - 5 pilots
  - + CESSDA, EOSC-NORDIC, DataverseNL
- Iterative mutual improvements ... ongoing process

**Best use formula: (F-UJI FAIR assessment + f2f FAIR consulting)**



# 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.

- Researcher
- Policy maker
- Research support (e.g. data steward, curator, data manager, librarian, information technology professional)
- Funder
- Publisher
- Other

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

- Research Infrastructure/eInfrastructure (e.g. data repository, service provider, library)
- University or Research Performing Organisation
- Research Performing Organisation
- Government
- eInfrastructure (e.g. repository or scientific data provider)
- Funding Body
- Publisher
- Industry
- Other

### 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