FAIRSFAIR
Fostering Fair Data Practices in Europe

Assessing the FAIRness of data holdings: Using F-UJI to make your repository more FAIR-enabling

Robert Huber
rhuber@uni-bremen.de
(on behalf of Task 4.5)

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

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

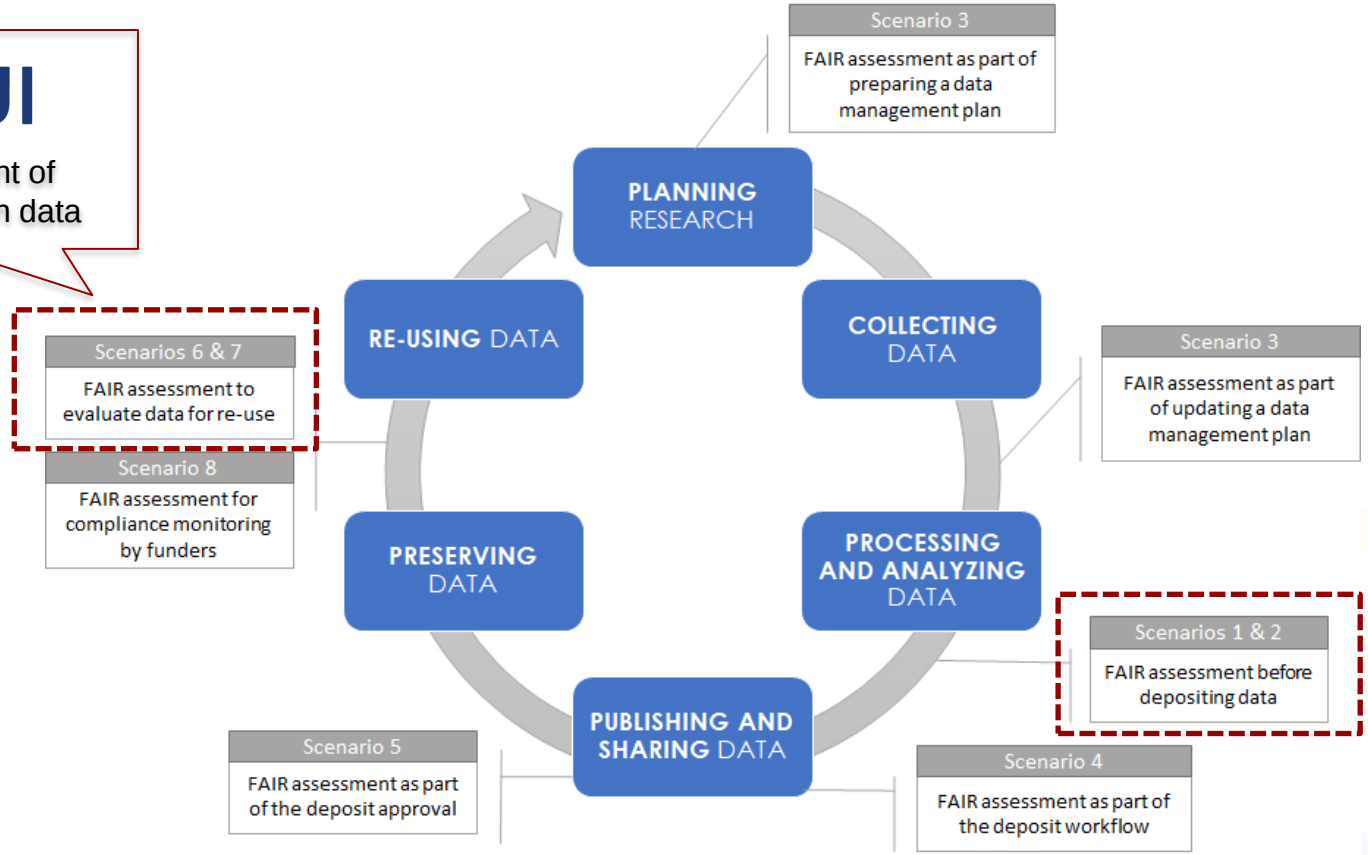
Rec. 25: Implement FAIR metrics to monitor uptake

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

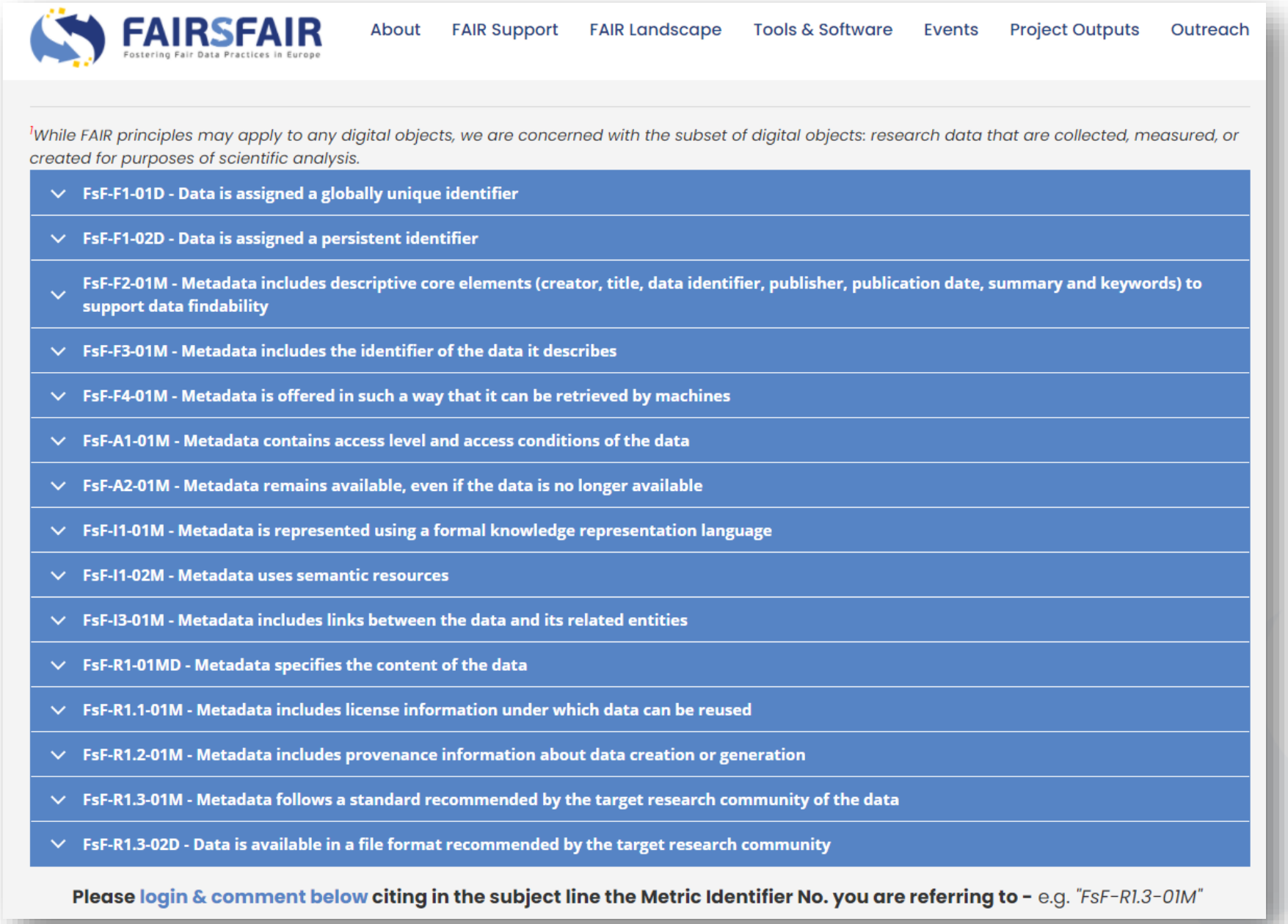
For more information, see D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories, https://doi.org/10.5281/zenodo.3678715

F-UJI
FAIR assessment of published research data



FAIR Aware
Assessing knowledge about FAIR

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

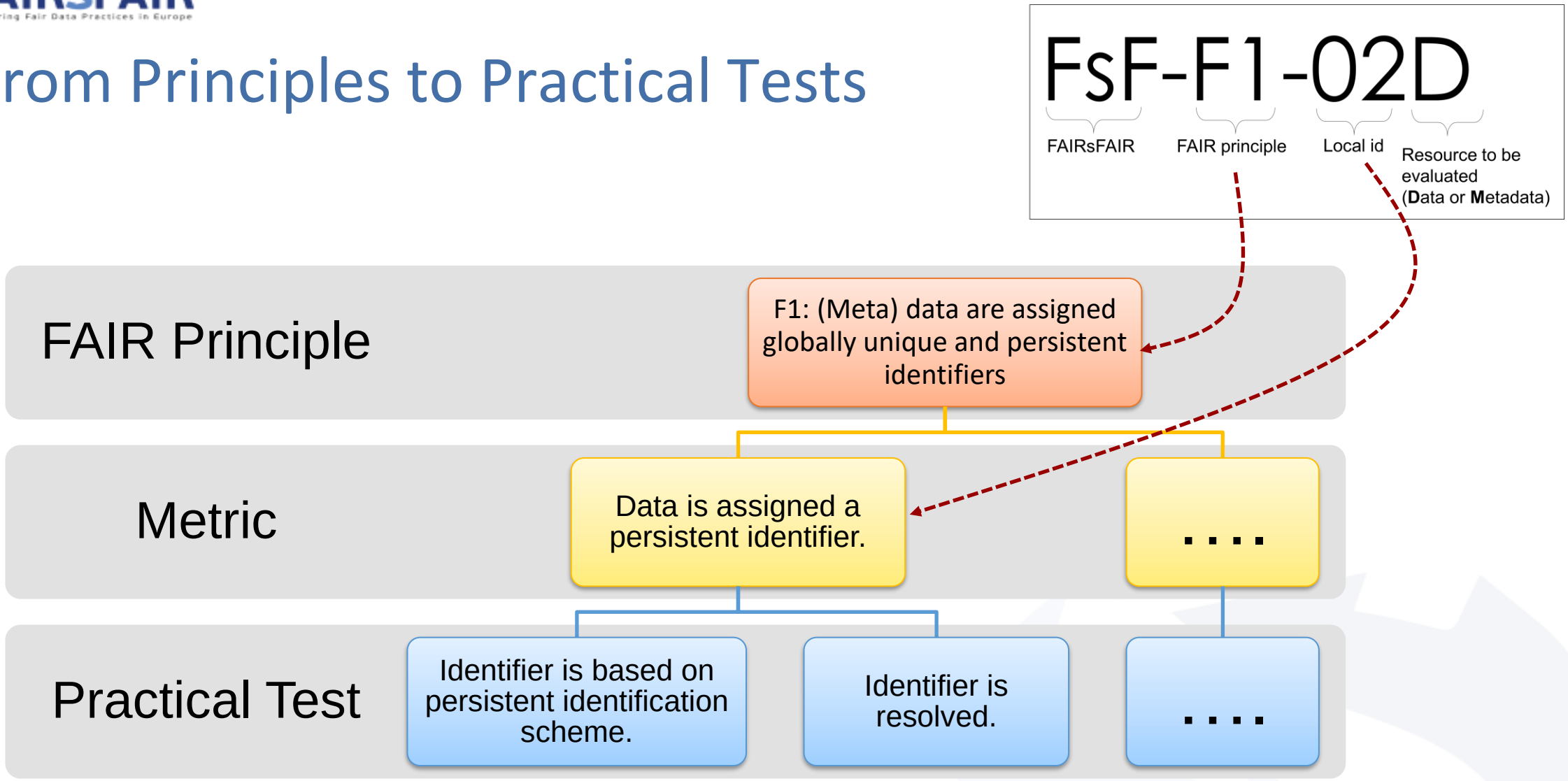


Object Assessment Metrics v0.4

We would love to hear your feedback!

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

From Principles to Practical Tests



From Principles to Practical Tests

Principle	Metrics	Practical Tests	Rationale
F1	**FsF-F1-02D** Data is assigned a persistent identifier	• 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.	• **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)

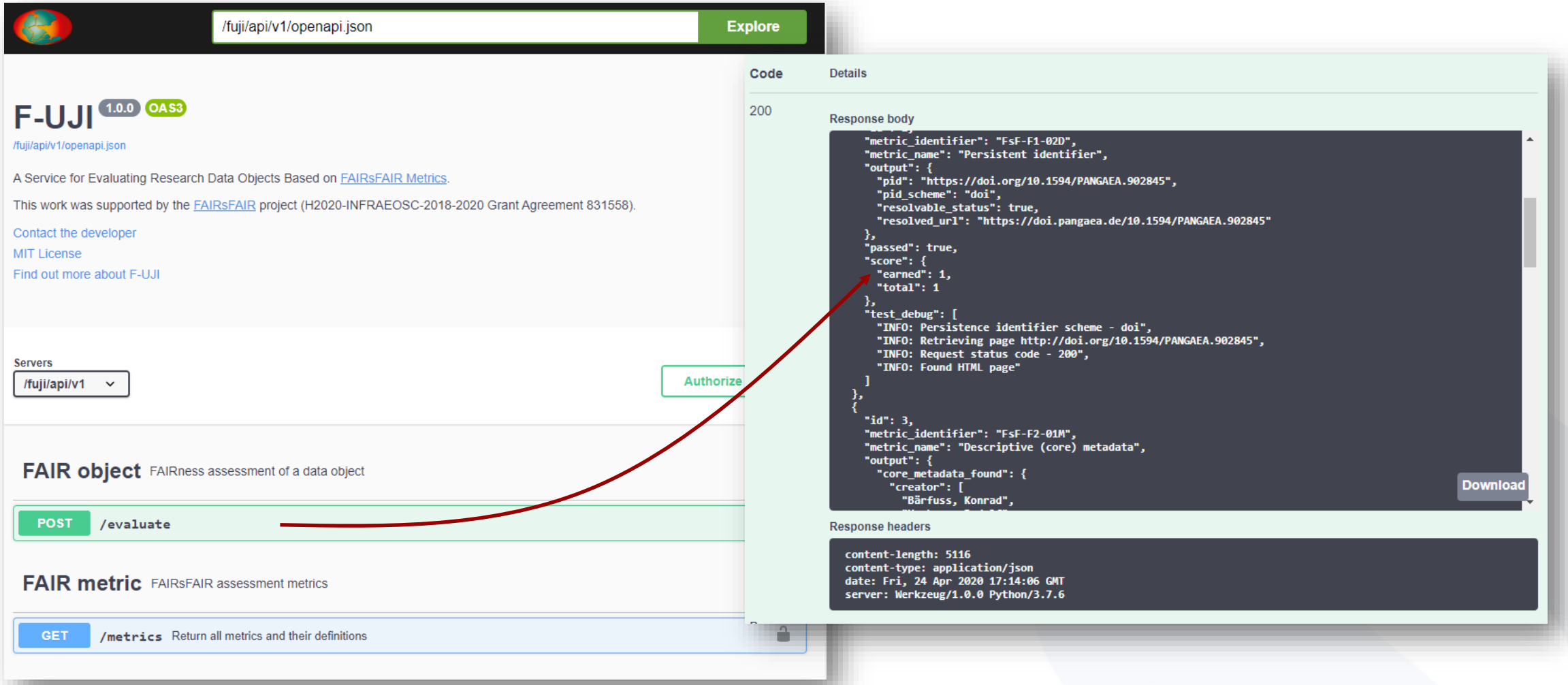
F2	**FsF-F2-01M**
	Metadata includes descriptive core elements to support data findability.
		• Some metadata (at all) has been made available via common (web) standards.
		• Minimum **core citation** metadata is specified (creator, title, publication date, publisher, and identifier)
		• Minimum **core descriptive** metadata is specified (creator, title, publisher, publication date, summary, keywords, identifier) through appropriate metadata fields.
			• **OAIS reference model** (ISO 14721:2012): 'Findable' => OAIS descriptive metadata
			• **Data citation**:
			  • *Recommendations* of Force11, ESIP, IASSIST, DataCite:
			• **Data description**:
			  • *Standards/Recommendations*: EOSC Datasets Minimum Information, DataCite Metadata Schema, W3C Recommendation Data on the Web, Data Catalog Vocabulary (DCAT-2)
			  • *Communality analysis* of common domain agnostic metadata standards

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

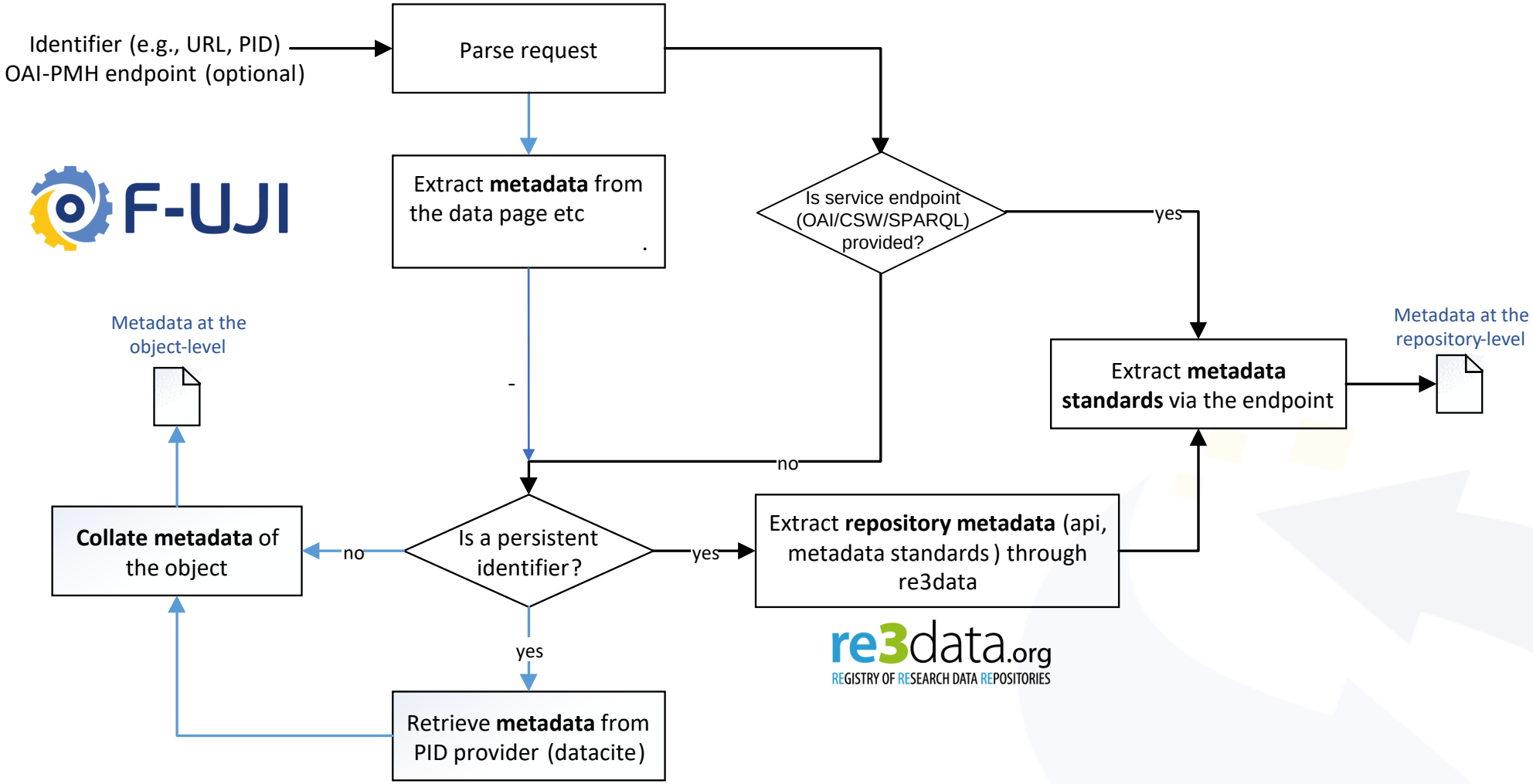
https://github.com/pangaea-data-publisher/fuji

https://www.f-uji.net

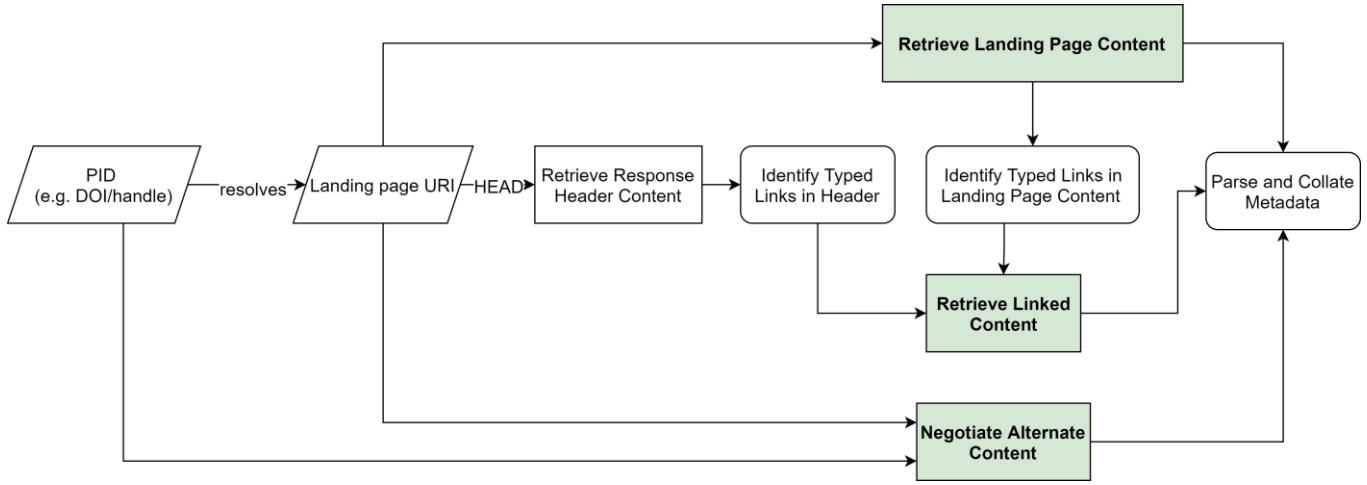
F-UJI – An Automated FAIR Data Assessment Tool



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	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

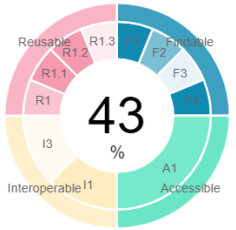
FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.

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.

Evaluation - Scoring

- Initially:
  - Pass/fail for each metric
  - Numerical score for each metric
  - Final FAIRness result:
    - # of passed;
    - total score (% of max)
- Now:
  - Maturity levels for each metric and principle
    - Based on CoreTrustSeal+FAIR Capability Maturity: 3 Tier model (https://doi.org/10.5281/zenodo.5346822)
  - Numerical score for each metric
  - Final FAIRness result:
    - Overall FAIR maturity level
    - total score (% of max)



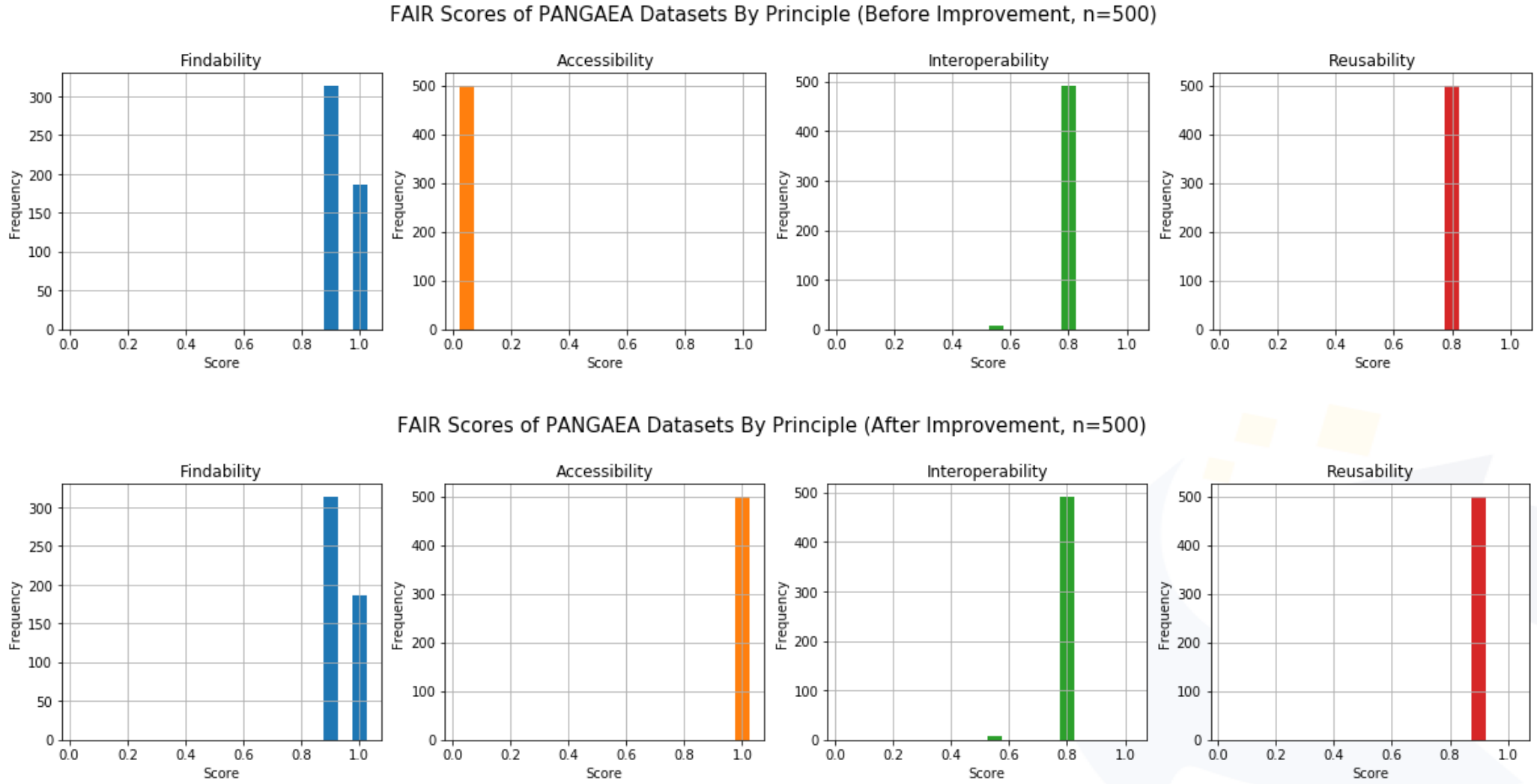
CTS FAIR CMM	F-UJI Maturity Levels
defined	advanced
managed	moderate
initial	initial
	incomplete

Testing & Consultation – the pilots

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

FAIRsFAIR Object Metric	Comment	Next Step
FsF-F1-01D Data is assigned a globally unique identifier.	OK	
FsF-F1-02D Data is assigned a persistent identifier.	OK	
FsF-F2-01M 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., https://doi.org/10.1594/PANGAEA.863175	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
FsF-F3-01M Metadata includes the identifier of the data it describes.	OK	
FsF-F4-01M Metadata is offered in such a way that it can be retrieved by machines.	OK	
FsF-A1-01M Metadata contains access level and access conditions of the data.	Total score = 1, Earned = 0 NO access information is available in metadata: **Possible ways to include access level metadata are:** **Schema.org** • if public use => isAccessibleForFree || free (Note: isAccessibleForFree Supersedes free)	Action: Uwe to include access level in the metadata (Done)

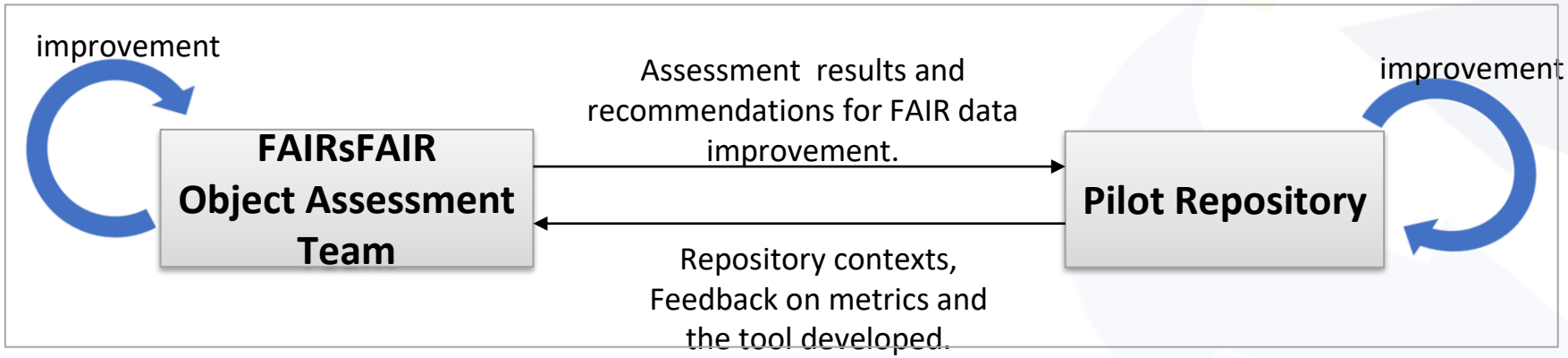
Before and After



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)



DEMO

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

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.

FAIRSFAIR
Fostering Fair Data Practices in Europe

Data Archiving and Networked Services
DANS

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 FAIRsFAIR 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 FAIRsFAIR Team (DANS, DCC, UniHB)

Find out more about FAIRsFAIR 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