FAIRification of data & services

EOSC Future with C-SCALE DICE EGI-ACE OpenAIRE NEXUS Reliance ARCHIVER

The EOSC Future, C-SCALE, DICE, EGI-ACE, OpenAIRE-Nexus, Reliance and ARCHIVER projects are funded by the European Union Horizon Programme calls INFRAEOSC-03-2020 and INFRAEOSC-07-2020 and H2020-ICT-2018-20

FAIRsFAIR's services framework

Patricia Herterich, DCC

FAIRsFAIR Framework for the assessment of FAIR-enabling services

Priority Recommendation

Rec. 13: "Develop metrics to certify FAIR services": More work is needed to extend the FAIR data principles for application to a wide range of data services, including registries, Data Management Planning tools, metadata standards and vocabulary bodies, identifier providers, software libraries and other cloud services…

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

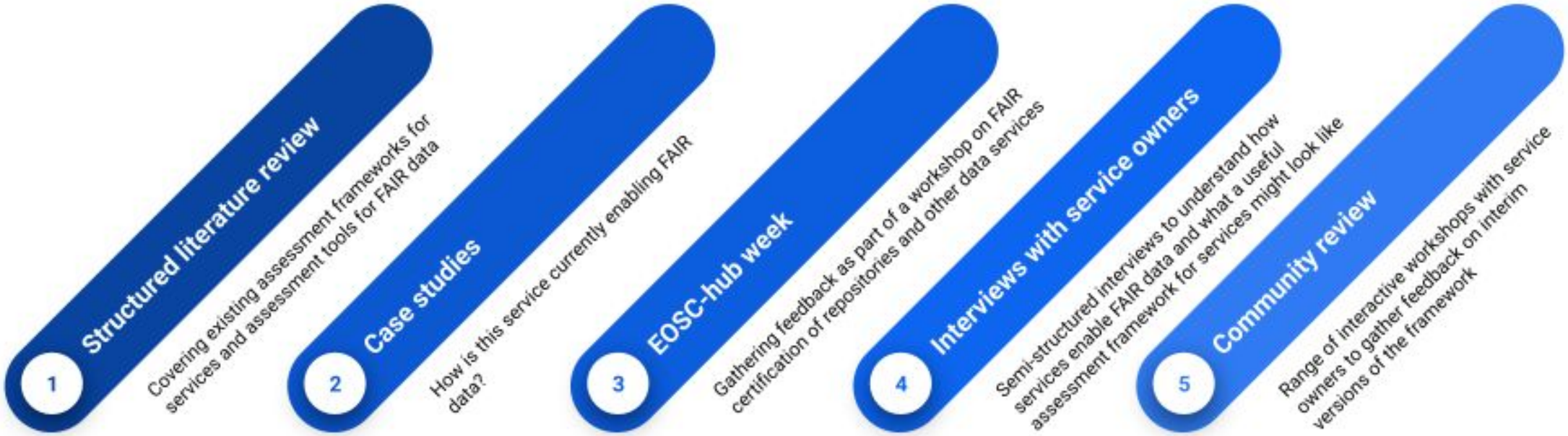
- Objective for FAIRsFAIR task on FAIR Services:

To deliver an assessment framework for data services that will help service owners to incrementally improve their services

→ stimulating an optimal interplay between digital objects and services

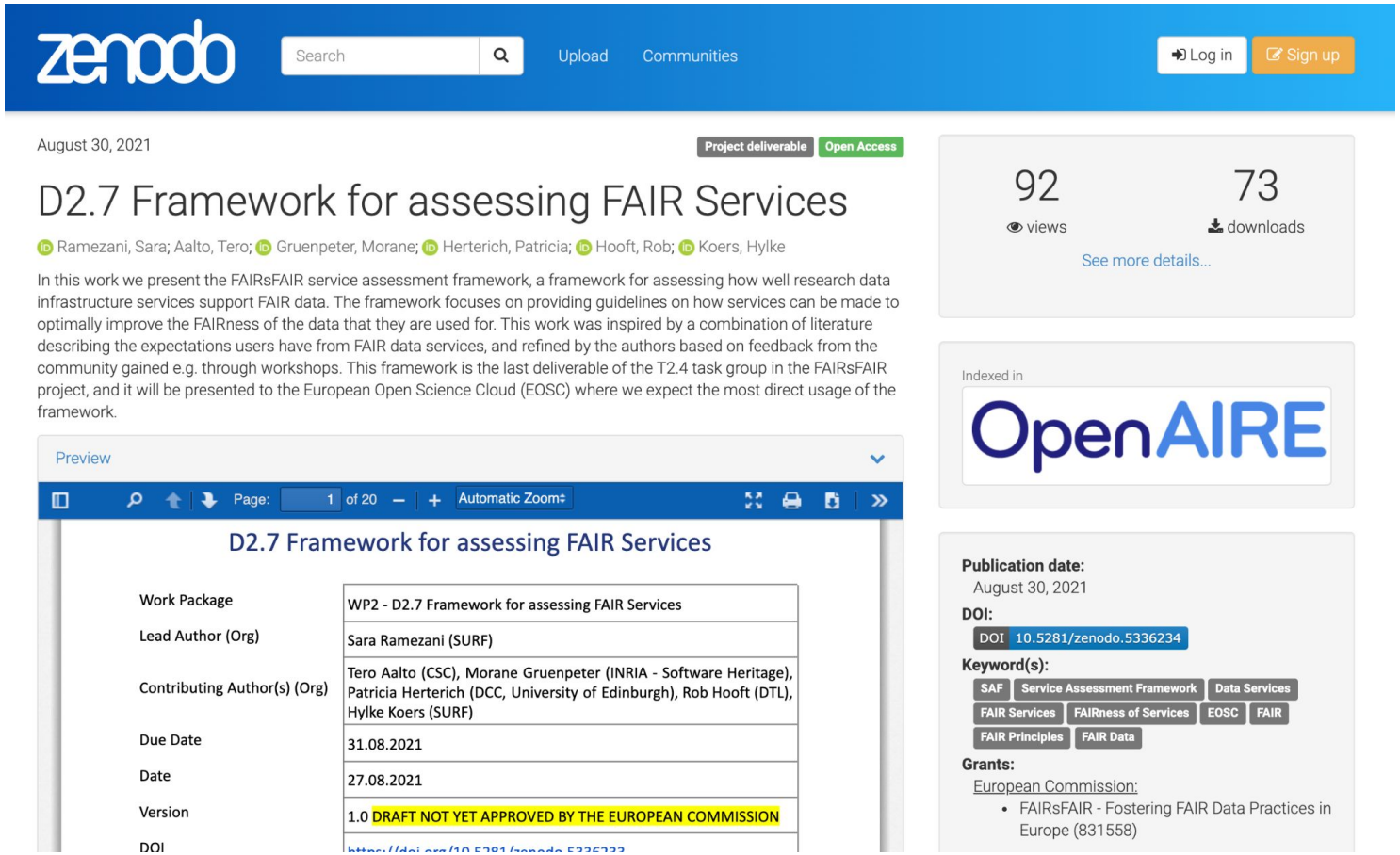
→ help realize the full potential of a truly FAIR ecosystem

From guiding principles to framework



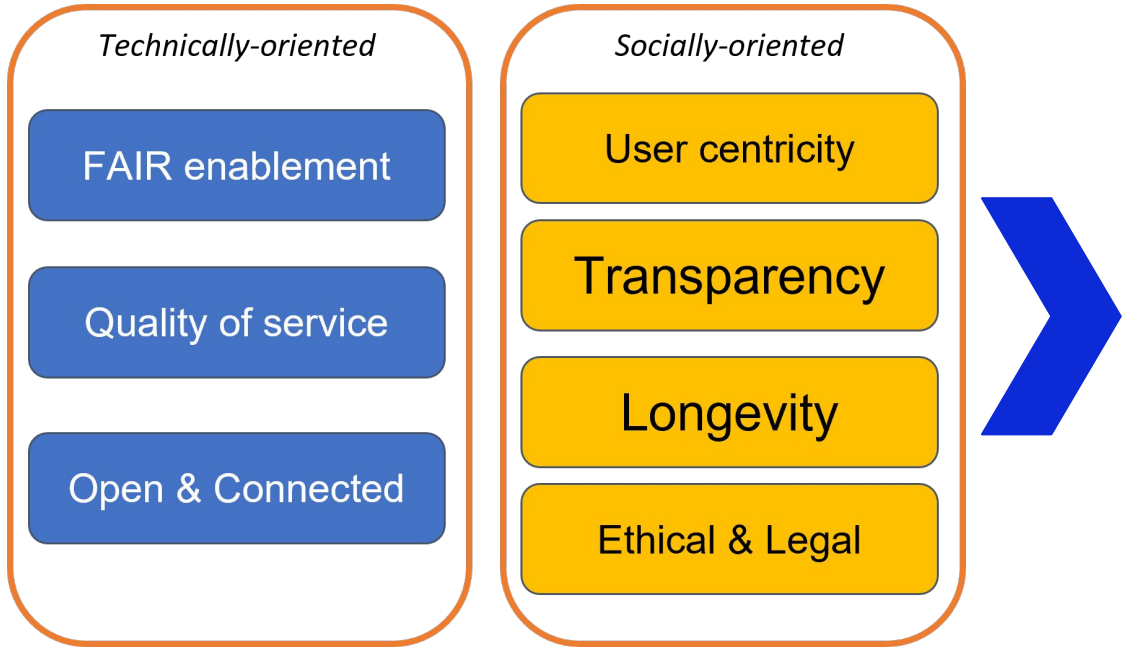
FAIRsFAIR service assessment framework

- 50 recommendations for service owners covering 7 different aspects



D2.7 Framework for assessing FAIR Services: https://doi.org/10.5281/zenodo.5336234

FAIRsFAIR service assessment framework



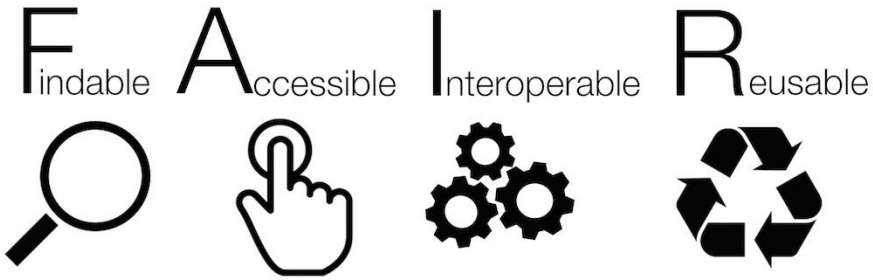
Aspect: SAF-O Open & Connected		
Objective: The service is operated in a low-barrier and inclusive way, seeking integrations and connections with other services and championing principles of openness consistent with Open Science and Open Research.		
Identifier	**Recommendation**	**Priority**
SAF-O-1	Publish clear, inclusive and non-discriminatory licences and/or terms of use. Enable wide access to the service.	Essential ☆☆☆
SAF-O-2	Seek integrations with other services rather than replicating functionalities, especially for common reusable infrastructure components. Provide documentation to ensure better sustainability for the network of integrations. Adopt the EOSC architectural components and standards as enablers for deep interoperability with other services in the EOSC portfolio[6].	Essential ☆☆☆

FAIRsFAIR service assessment framework - Aspect: FAIR enablement

Objective: The service enables FAIR data by elevating the FAIRness of digital objects and/or supporting the FAIRification process. FAIR enablement is actively driven through the implementation of community-supported standards and interoperability frameworks.

The following classification defines the levels in which a service affects the FAIRness of data on which it operates:

- ***Enable:***
  - ***Augment****:* The service provides elements improving FAIRness of the digital object for example automatically assigning a PID;
  - ***Facilitate***: The service actively helps to realize a particular FAIR principle — for example by allowing the user to add metadata or enabling discoverability;
- ***Respect***: The service neither actively enables a particular FAIR principle nor interferes with it — it can be said to respect the "FAIR-in-FAIR-out" principle;
- ***Reduce***: The service actually makes data less FAIR — at least for a particular principle — for example by detaching metadata or a PID when it acts on a digital object.



6 recommendations:

- 4 Essential
- 1 Important
- 1 Useful

FAIRsFAIR service assessment framework - Technically-oriented aspects

Quality of service

Objective:

The service is delivered in a reliable, secure, high-quality way, consistent with its specifications.

11 recommendations:

- 3 Essential
- 5 Important
- 3 Useful

Open & Connected

Objective:

The service is operated in a low-barrier and inclusive way, seeking integrations and connections with other services and championing principles of openness consistent with Open Science and Open Research.

9 recommendations:

- 4 Essential
- 5 Important

FAIRsFAIR service assessment framework - Socially-oriented aspects

User centricity

Objective:

The service is managed so that it serves the (possibly evolving) goals of the user community and maximises usability while minimizing burden.

8 recommendations:

- 4 Essential
- 2 Important
- 2 Useful

Transparency

Objective:

The service provider communicates with its stakeholders in a transparent manner.

6 recommendations:

- 2 Essential
- 2 Important
- 2 Useful

Longevity

Objective:

The service provider designs the service with a timeframe for the maintenance and sustainability of the service in mind and implements measures accordingly, considering the researchers' need for reproducible research.

4 recommendations:

- 2 Essential
- 2 Important
- 0 Useful

Ethical & Legal

Objective:

The service complies with all applicable legal and ethical guidelines, in a transparent and auditable way.

6 recommendations:

- 3 Essential
- 2 Important
- 1 Useful

Using the framework

The framework is useful to

- Check your own documentation and processes
- Getting better at becoming FAIR-enabling: it's for self-improvement

The framework is not

- For externally assessing a service
- Machine-actionable
- Showing off

F-UJI Tool

Robert Huber, PANGAEA

F-UJI – A FAIR data assessment tool

F-UJI

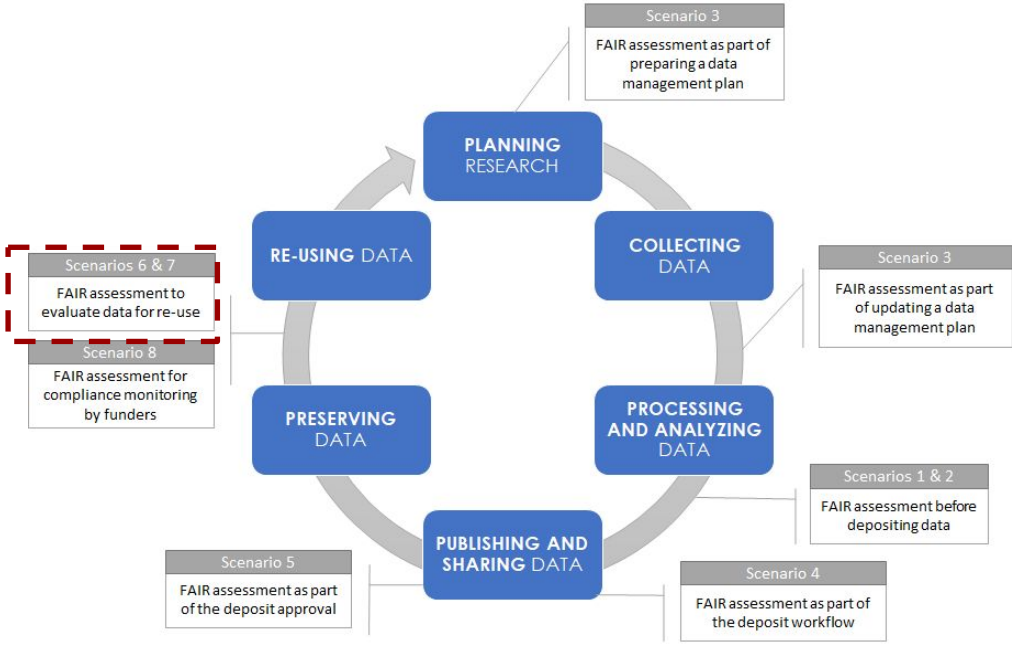
FAIRsFAIR Task 4.5: Pilot FAIR assessment

Two main components – assessment metrics and tool.

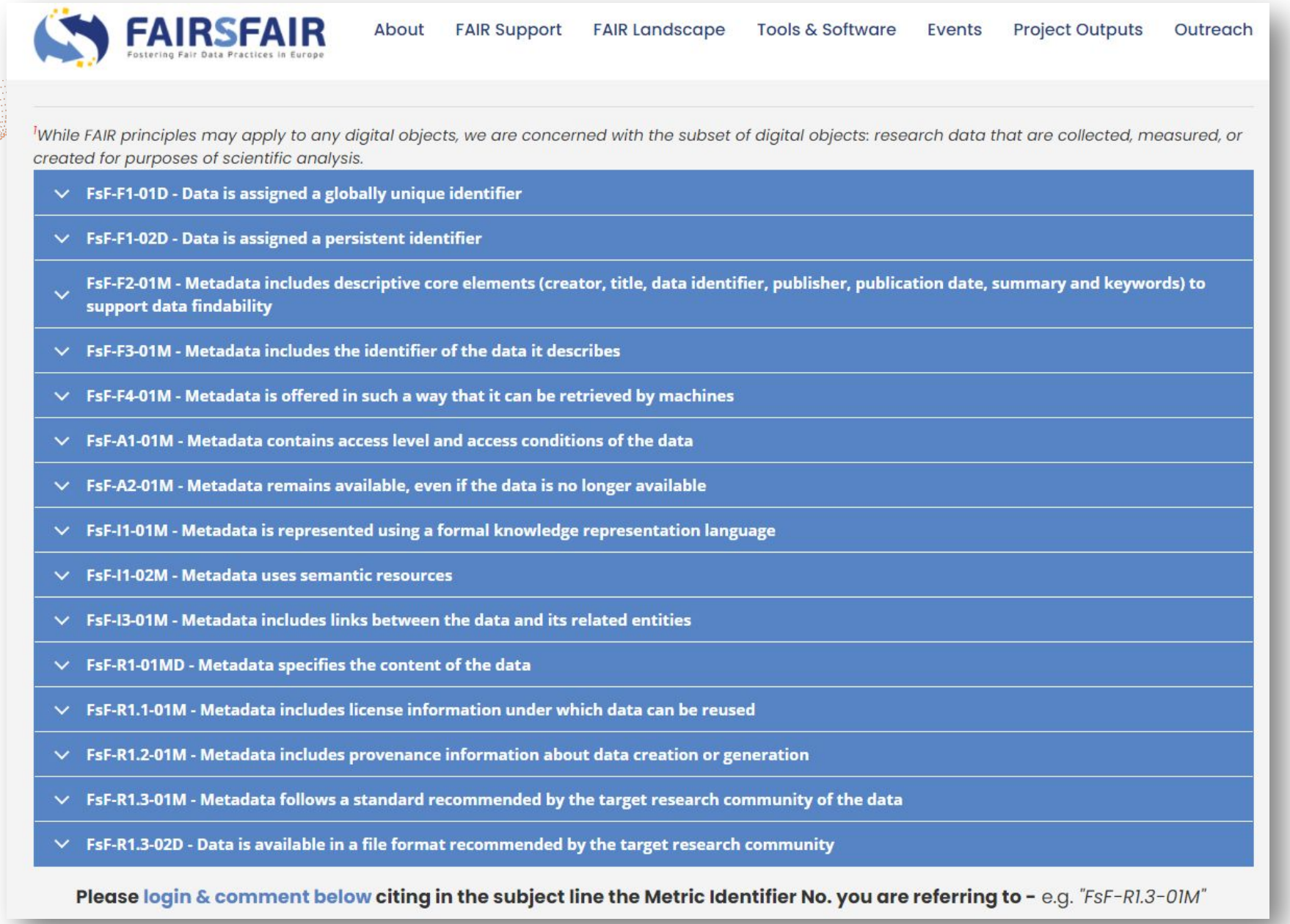
European Commission Expert Group on FAIR Data. 2018. 'Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data.' https://doi.org/10.2777/1524

<u>Priority Recommendations</u>
Rec. 8: Facilitate automated processing
Rec. 12: Develop metrics for FAIR Digital Objects

<u>Supporting Recommendations</u>
Rec. 25: Implement FAIR metrics to monitor uptake



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



FAIRSFAIR
Fostering Fair Data Practices in Europe

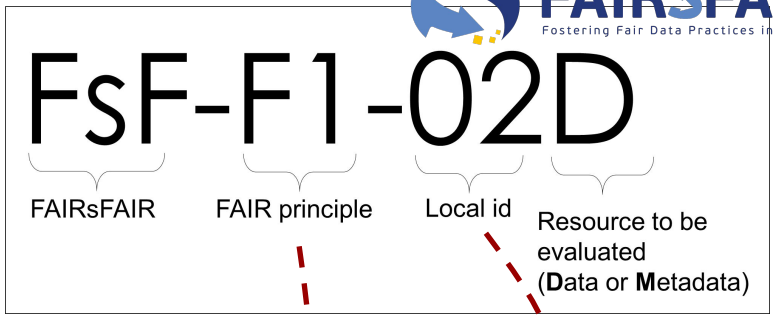
Object Assessment Metrics v0.5

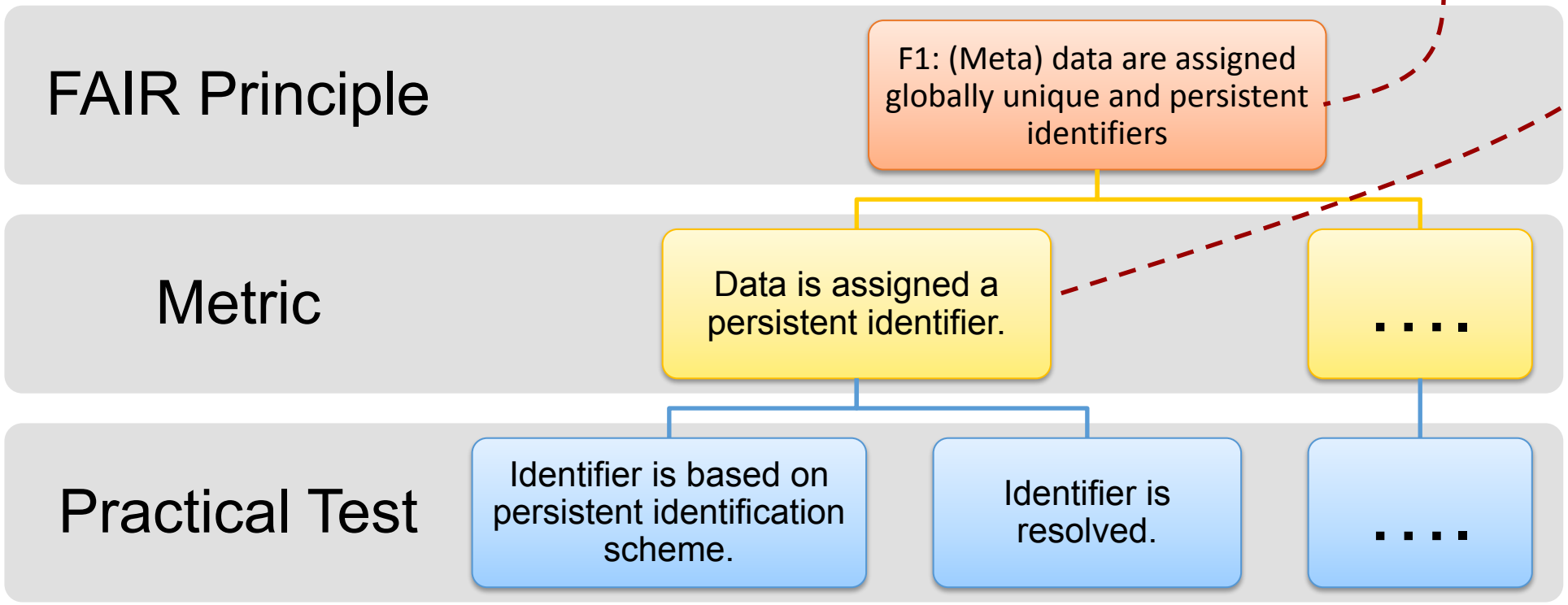
DOI 10.5281/zenodo.6461229

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

F-UJI – A FAIR assessment tool

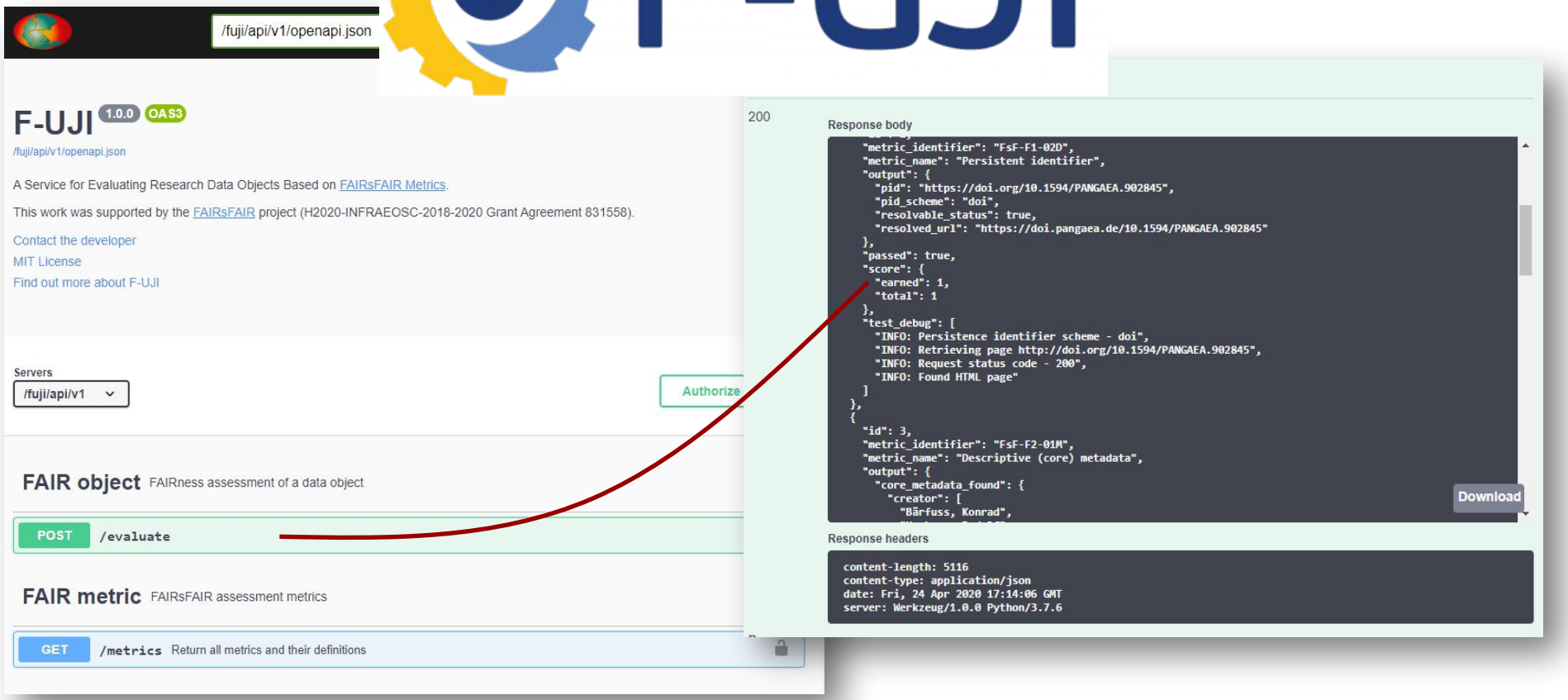
From Principles to Practical Tests





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

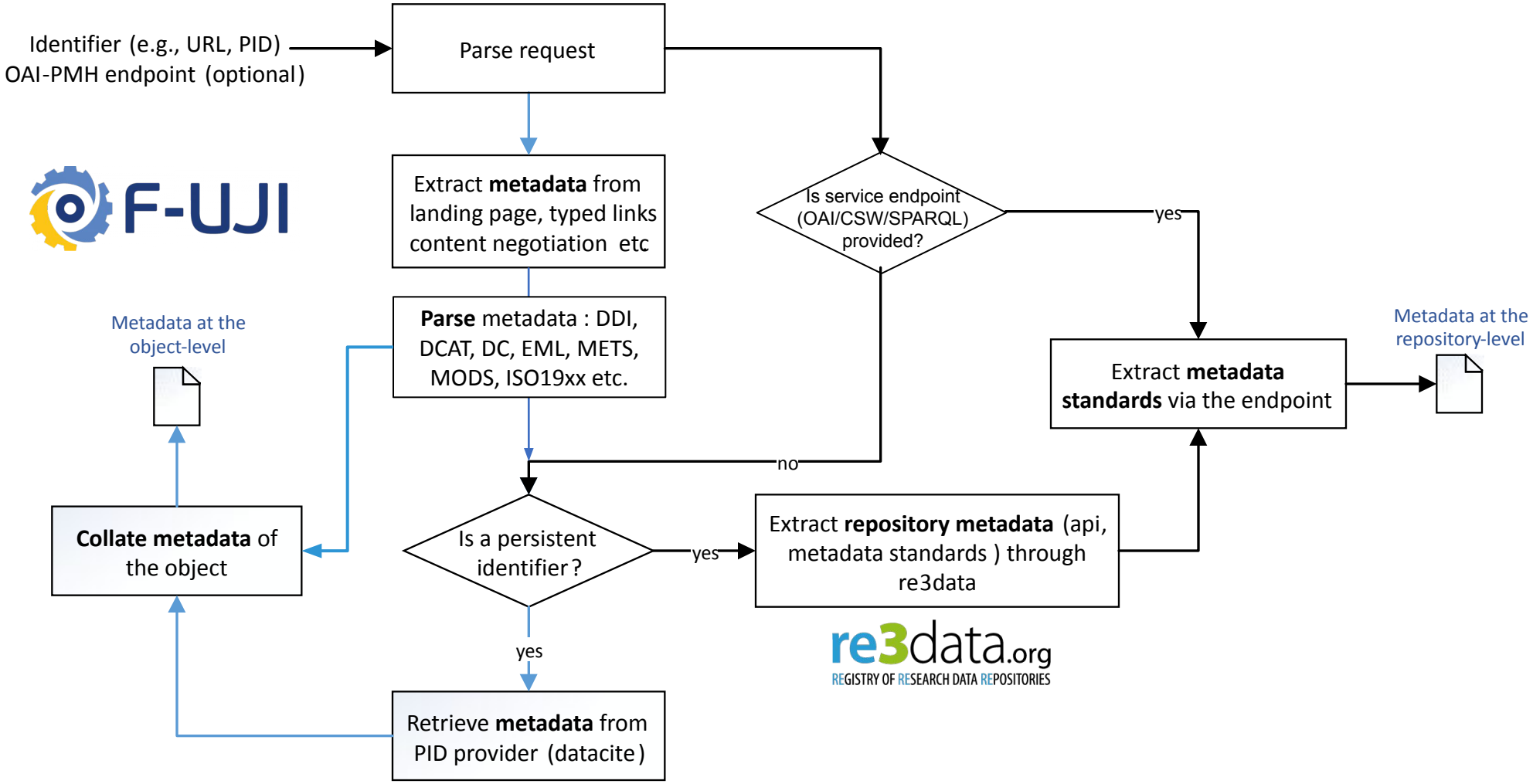
F-UJI



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

https://www.f-uji.net

High Level Flow ([Meta-]data Gathering)



F-UJI for humans - web client demo

Assessment Results:

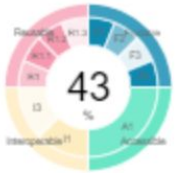
Evaluated Resource:

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.

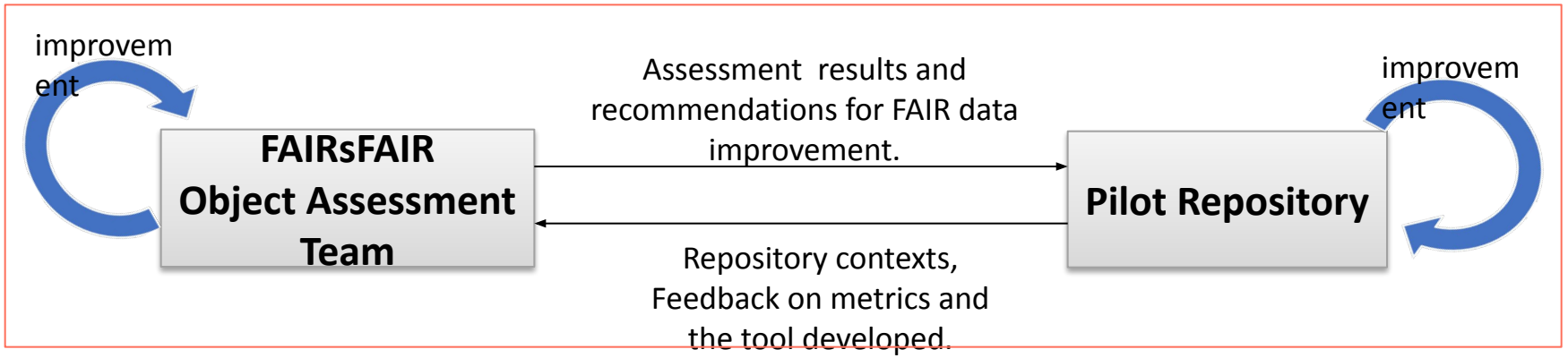
FsF-F3-01M - Metadata includes the identifier of the data it describes.

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

Lessons learned

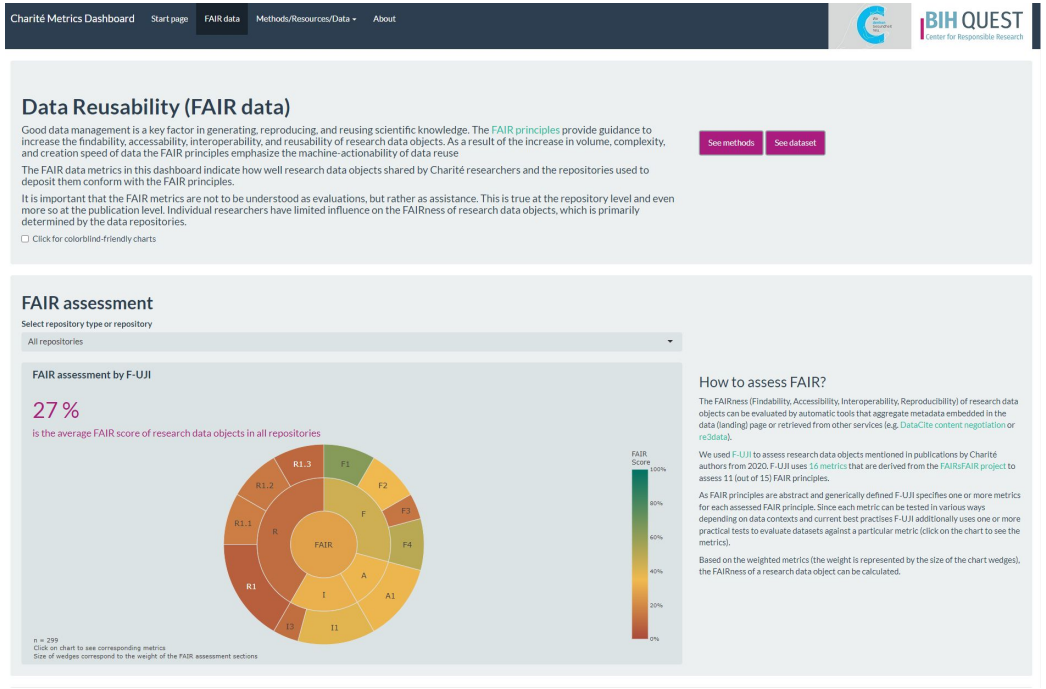
- Automatised FAIR assessment of research data objects is possible
- Supporting a very large and diverse community
  - 5 pilots
  - + CESSDA, EOSC-NORDIC, DataverseNL
- Iterative mutual improvements … ongoing process

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



F-UJI Uptake

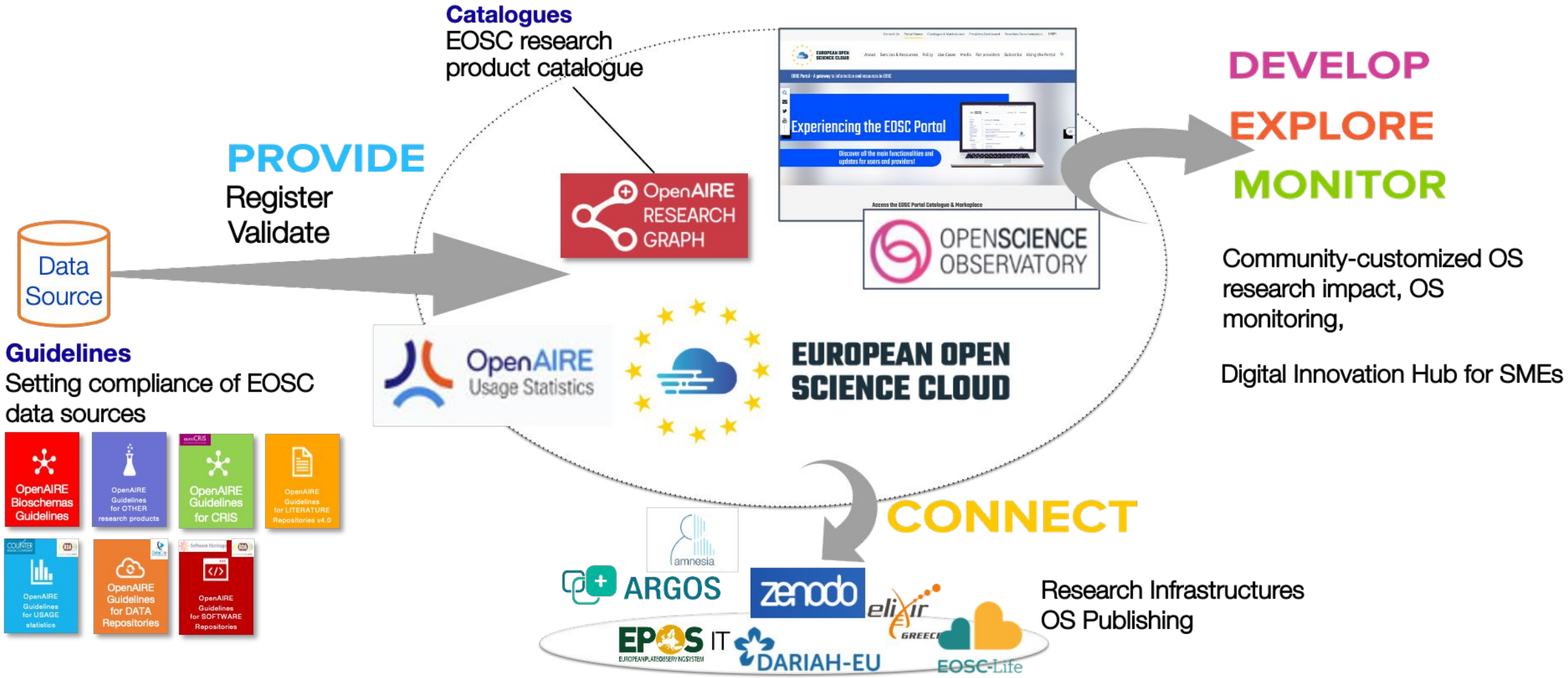
- Open Source community:
  - 12 contributers, 18 forks, 3 clients: (R, web)
- Dataset assessments:
  - ~10.000 individual tests via f-uji.net
  - > n-thousands during repo tests (see below)
- Repository assessments:
  - FAIRsFAIR pilots: 5+3 repos assessed
  - DANS DGRTD project
  - Institutional tests (e.g. Charité Berlin, UVP, Novartis)
- Publications:
  - Peters-von Gehlen et al. (2022). Recommendations for Discipline-Specific FAIRness Evaluation Derived from Applying an Ensemble of Evaluation DOI: 10.5334/dsj-2022-007
  - Mai (2022). Transparency of Covid-19-related Research in Dental Journals. DOI: 10.3389/froh.2022.871033
  - Huber (2021). Integrating Data and Analysis Technologies within Leading Environmental Research Infrastructures: Challenges and Approaches. DOI: 10.1016/j.ecoinf.2021.101245



OpenAIRE Validator

Paolo Manghi, OpenAIRE

OpenAIRE and EOSC



EXPLORE | PROVIDE | CONNECT | MONITOR | DEVELOP

OpenAIRE | PROVIDE

HOME ABOUT SIGN IN

Your data is valuable. Get connected. Participate

The Provide Dashboard is a one-stop-service where content providers interact with OpenAIRE and become a building block of a global Open Research community. A gateway to the European Open Science Cloud.

SIGN IN

OpenAIRE | PROVIDE

- R Register
- V Validator

ADMIN

- H Help Texts
- R Registrations
- M Metrics

Hide Controls

You have not yet registered any repositories

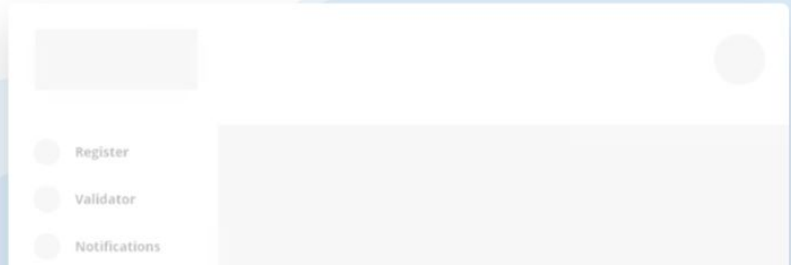
Register

Register data sources in the OpenAIRE infrastructure.

Validate

Validate data sources against OpenAIRE guidelines.

Take a look at the Provide Dashboard and functionalities



Enrich

Improve your metadata. Get more connections.

Our newly released OA Broker service offers a wealth of informa

OpenAIRE | PROVIDE

PM

- R Register
- V Validator
  - Validate
  - Validation History

ADMIN

- H Help Texts
- R Registrations
- M Metrics

Validate your datasource

Literature repository

Data repository

CRIS systems

BETA

FAIR

FAIR assessment

OpenAIRE Guidelines for Literature Repository Managers v4.0

- Based on established standards: Dublin Core and DataCite metadata scheme
- For description of textual and data publications
- Application Profile
- Controlled Vocabularies aligned with the other OpenAIRE Guidelines

http://dx.doi.org/10.5281/zenodo.1299203
(Released Nov-2018)

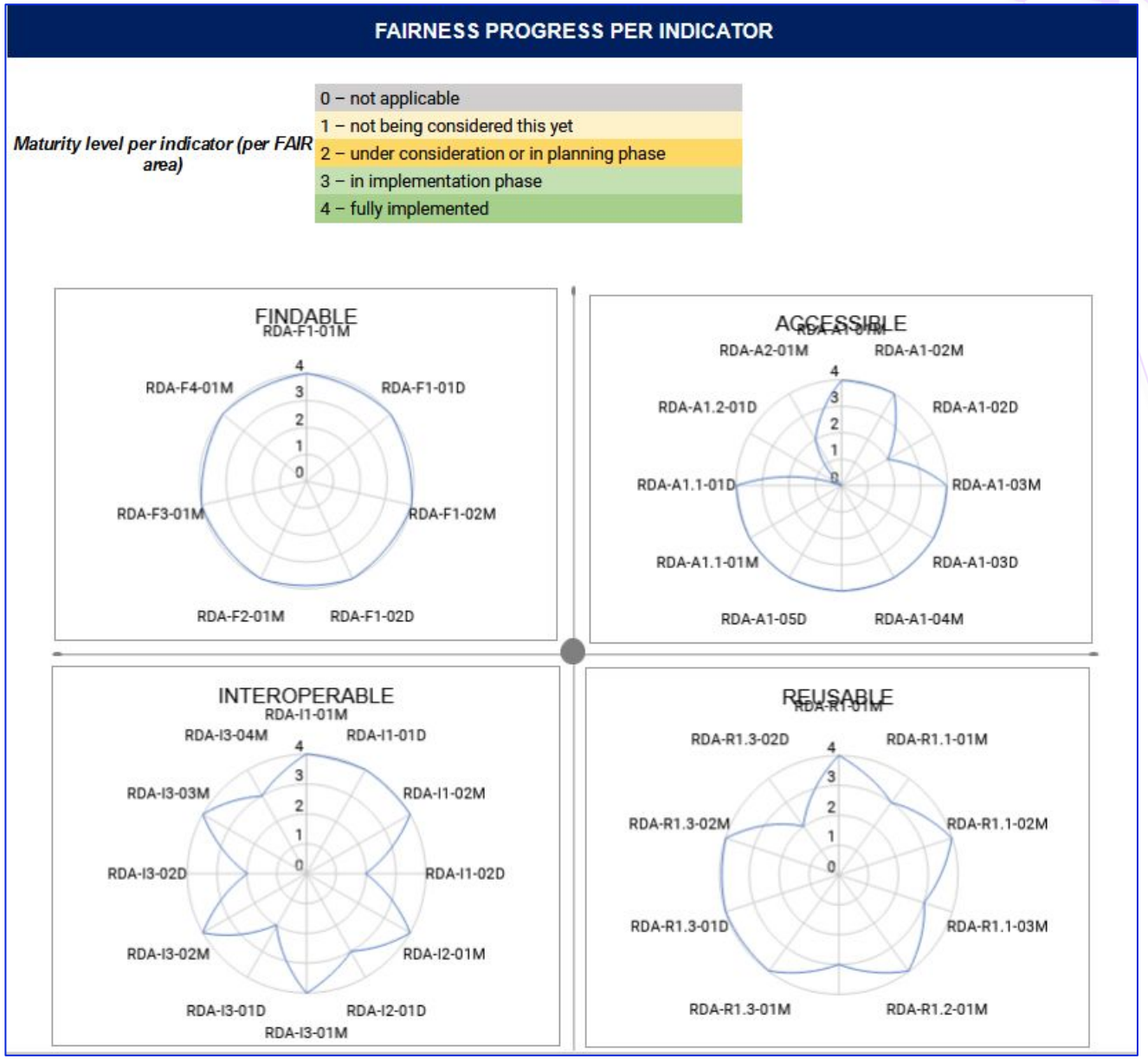
OpenAIRE-Field	Metadata Element	Refinement by Vocabulary
Title (M)	datacite:title	title type
Creator (M)	datacite:creator	name type
Contributor (MA)	datacite:contributor	name type contributor type
Funding Reference (MA)	oaire:fundingReference	funderIdentifier type
Alternate Identifier (R)	datacite:alternateIdentifier	alternateIdentifier type
Related Identifier (R)	datacite:relatedIdentifier	relatedIdentifier type relation type resourcetype general
Embargo Period Date (MA)	datacite:date	date type
Language (MA)	dc:language	IETF BCP 47, ISO 639-3
Publisher (MA)	dc:publisher	
Publication Date (M)	datacite:date	date type
Resource Type (M)	oaire:resourceType	COAR Resource Type Vocabulary
Description (MA)	dc:description	
Format (R)	dc:format	
Resource Identifier (M)	datacite:identifier	identifier type
Access Rights (M)	datacite:rights	COAR Access Right Vocabulary
Source (R)	dc:source	
Subject (MA)	datacite:subject	
License Condition (R)	oaire:licenseCondition	
Coverage (R)	dc:coverage	
Size (O)	datacite:size	
Geo Location (O)	datacite:geoLocation	
Resource Version (R)	oaire:version	COAR Version Vocabulary
File Location (MA)	oaire:file	
Citation Title (R)	oaire:citationTitle	
Citation Volume (R)	oaire:citationVolume	
Citation Issue (R)	oaire:citationIssue	
Citation Start Page (R)	oaire:citationStartPage	
Citation End Page (R)	oaire:citationEndPage	
Citation Edition (R)	oaire:citationEdition	
Citation Conference Place (R)	oaire:citationConferencePlace	
Citation Conference Date (R)	oaire:citationConferenceDate	
Audience (O)	dcterms:audience	

OpenAIRE NEXUS

FAIRification

based on Research Data Alliance - FAIR Data Maturity Model: specification and guidelines [https://doi.org/10.15497/RDA00050]

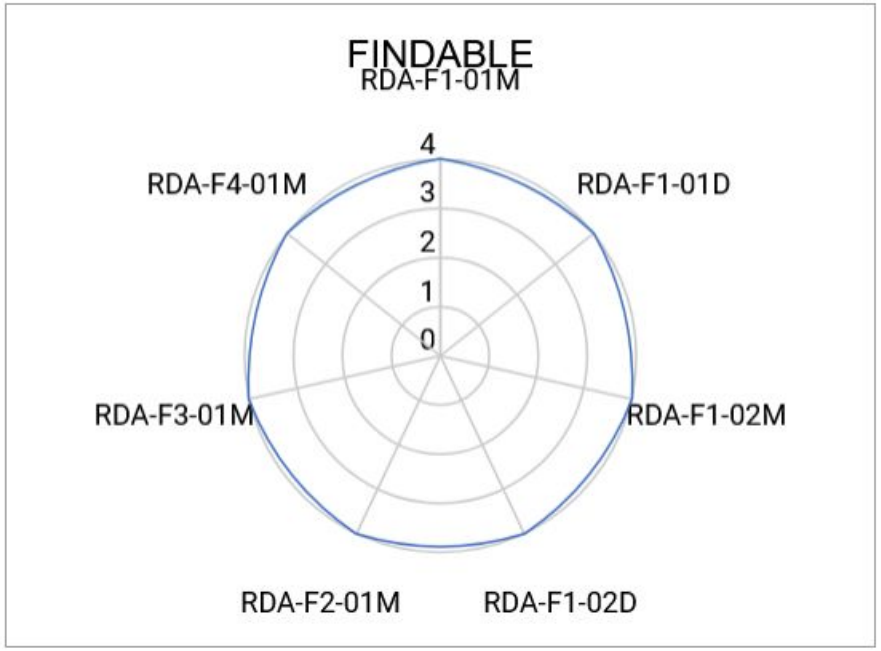
- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**euseable



FAIRification

- Findable

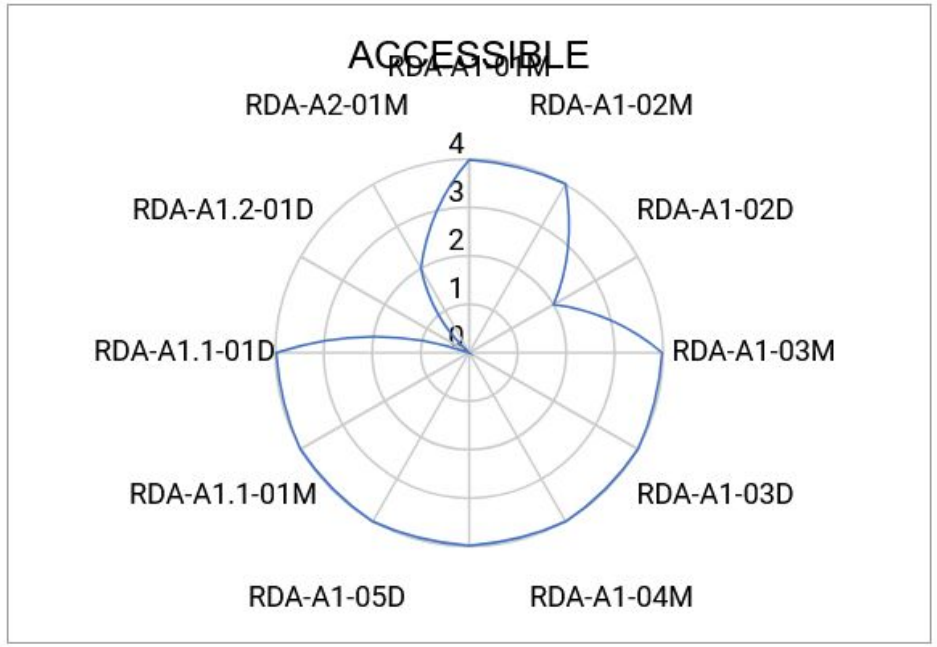
Indicator Id	FAIR Indicator	Priority	met in the guidelines	Primarily Supported in the Guidelines
RDA-F1-01M	Metadata is identified by a persistent identifier	Essential	y	Resource Identifier
RDA-F1-01D	Data is identified by a persistent identifier	Essential	y	Resource Identifier
RDA-F1-02M	Metadata is identified by a globally unique identifier	Essential	y	Resource Identifier
RDA-F1-02D	Data is identified by a globally unique identifier	Essential	y	Resource Identifier
RDA-F2-01M	Rich metadata is provided to allow discovery	Essential	y	implicitly
RDA-F3-01M	Metadata includes the identifier for the data	Essential	y	File Location
RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	Essential	y	implicitly



FAIRification

- Accessible

A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	Important	y	Access Rights
A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human interventio n)	Essential	y	not controlled by / in the scope of the Guidelines
A1	RDA-A1-02D	Data can be accessed manually (i.e. with human interventio n)	Essential	n	not controlled by / in the scope of the Guidelines
A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	Essential	y	implicitly supported by the OAI-PMH protocol but also digital objects minted with a DOI
A1	RDA-A1-03D	Data identifier resolves to a digital object	Essential	y	Resource Identifier
A1	RDA-A1-04M	Metadata			

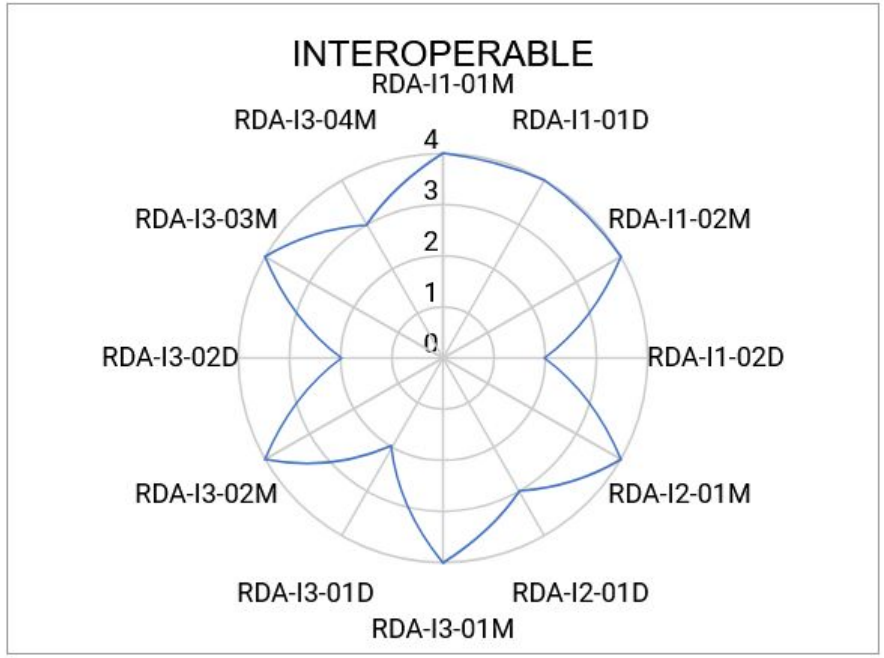


RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	Useful

FAIRification

- Interoperable

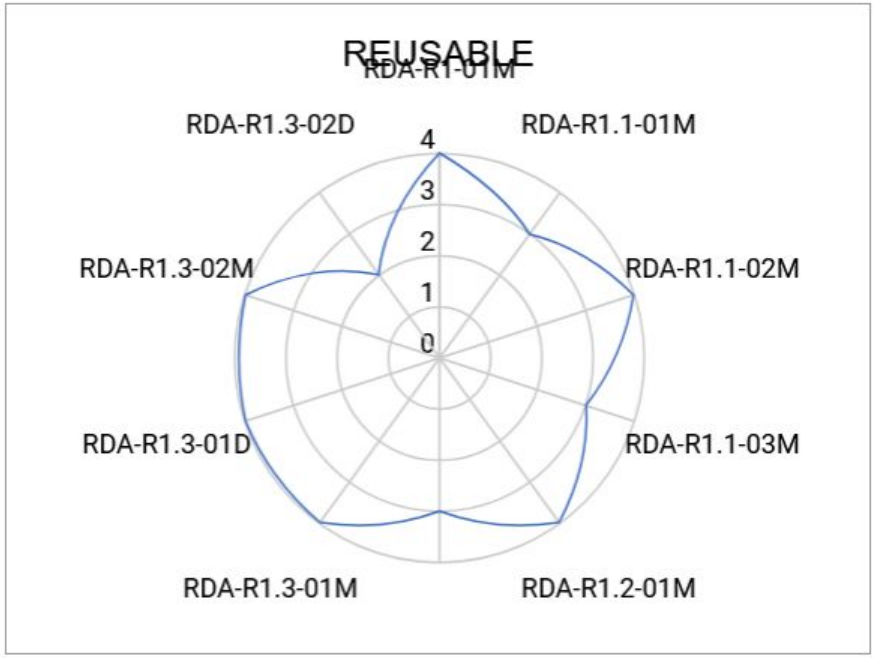
I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	Important	y	Subject
I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	Important	y	Format
I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	Important	y	knowledge representation is cross-domain and on the level of XML
I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	Important	n	not explicitly supported
I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	Important	y	supported for some vocabularies in the Guildeines, e.g. COAR vocabularies, CC-licenses vocabularies
I2	RDA-I2-01D	Data uses FAIR-compliant	Useful		not supported by the



FAIRification

- Reusable

R1	RDA-R1-01M	Plurality of accurate and relevant attributes are provided to allow reuse	Essential	y	yes, as long as it concerns bibliographic information
R1.1	RDA-R1.1-01M	Metadata includes information about the licence under which the data can be reused	Essential	y	License Condition
R1.1	RDA-R1.1-02M	Metadata refers to a standard reuse licence	Important	y	License Condition
R1.1	RDA-R1.1-03M	Metadata refers to a machine-under standable reuse licence	Important	y	License Condition
R1.2	RDA-R1.2-01M	Metadata includes provenance information according to community-specific standards	Important	n	not supported
R1.2	RDA-R1.2-02M	Metadata includes provenance information according to a cross-commu	Useful		



References

- latest version on Guidelines v4.1-rc1 at https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/latest/

- Schema and examples on github https://github.com/openaire/guidelines-literature-repositories

Feedback channels

- Making annotations on the latest guidelines pages

- Creating Issue at GitHub https://github.com/openaire/guidelines-literature-repositories/issues

- gDocument (without credentials) https://docs.google.com/document/d/1aFWrkBO_f_GTWWTAfco4uoUphEbtOOdJVD_2ohYw7Ek/edit#