FAIRSFAIR

Fostering Fair Data Practices in Europe

Making your repository more FAIR-enabling

December 2, 2020 - CODATA & GFISCO FAIR Convergence Symposium

FAIRsFAIR project: Elizabeth Newbold (STFC), Joy Davidson (DCC), Claudia Behnke (SURF), Robert Huber (UniHB/PANGAEA/MARUM), Linas Čepinskas and Frans Huigen (DANS)

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

- **Introduction** - Frans Huigen (10 minutes)
- **FAIRsFAIR Certification Support Programme** - Frans Huigen
- **Building capacity in the Community** - Elizabeth Newbold & Joy Davidson
- **Implementing a FAIR Data Point** - Claudia Behnke
- **Supporting the assessment of FAIR data in trustworthy repositories: a demonstration of the FAIR-Aware and F-UJI tools**
  Linas Cepinskas and Robert Huber
- **Questions & exchange of ideas**

Introducing...

Today's goal: share some of the ongoing work of the FAIRsFAIR project to

- improve the trustworthiness of repositories through CoreTrustSeal certification and FAIR alignment
- improve the interoperability of repositories
- increase data stewardship skills and capacity
- assess FAIR data in trustworthy repositories

FAIRSFAIR
Fostering Fair Data Practices in Europe

FAIRsFAIR partners

Data Archiving and Networked Services
DANS

Trust-IT Services
Communicating ICT to markets

Universiteit van Amsterdam

DCC

Helsingin yliopisto
Helsingfors universitet
University of Helsinki

CODATA
Committee on Data
International Science Council

eua
European University Association

uc3m | Universidad Carlos III de Madrid

CSC

Inria

Universidade do Minho

SURF SARA

SPARC Europe

University of Essex

DataCite
FIND, ACCESS, AND REUSE DATA

Science & Technology Facilities Council | UK Research and Innovation

eSciencedata Factory

Georg-August-Universität Göttingen

INES
Centre Informatique National de l'Enseignement Supérieur

RDA
Research Data Alliance

DTL

Universität Bremen

FAIRsFAIR in a nutshell

Call: H2020-INFRAEOSC-5c

Budget: €10 million

Length: 36 months
Starting date: March 1 2019

22 partners from 8 EU members
6 core partners

Project objective

Help **survey the landscape of FAIR activities** in relation to the EOSC and identify where dialogue and collaboration can be encouraged.

Create a **basis for harmonisation** efforts to bring together the various actors working in the **FAIR ecosystem** and build a **functioning EOSC** and **active community** around EOSC.

FAIRSFAIR

Fostering Fair Data Practices in Europe

MAIN OUTPUTS

March 2019 - February 2022

DATA PRACTICES

- **Reports**
  - FAIR requirements for persistence and interoperability
  - Guidelines for ontology design and vocabulary interoperability
  - Basic framework for services enabling FAIR (including software)
- **Solutions for interoperability and machine accessibility for FAIR-aligned repositories**
- **Prototype for interoperability of repositories**
- **Workshops and hackathons: Recommendations for FAIR Semantics and Semantics in FAIR**

DATA POLICY

- **Reports**
  - Recommendations on data policy and analysis of practice
  - Integration of meta-data catalogues
  - White paper on alignment and synchronisation around FAIR, Open Science and EOSC
- **Support programme for repositories to reach FAIR compliance**

CERTIFICATION

- **European network of trustworthy repositories enabling FAIR data**
- **Support and guidance for certification of data repositories**
- **Tool to identify relevant trustworthy certified repositories**
- **Pilots to support the assessment of FAIR data in trustworthy repositories**

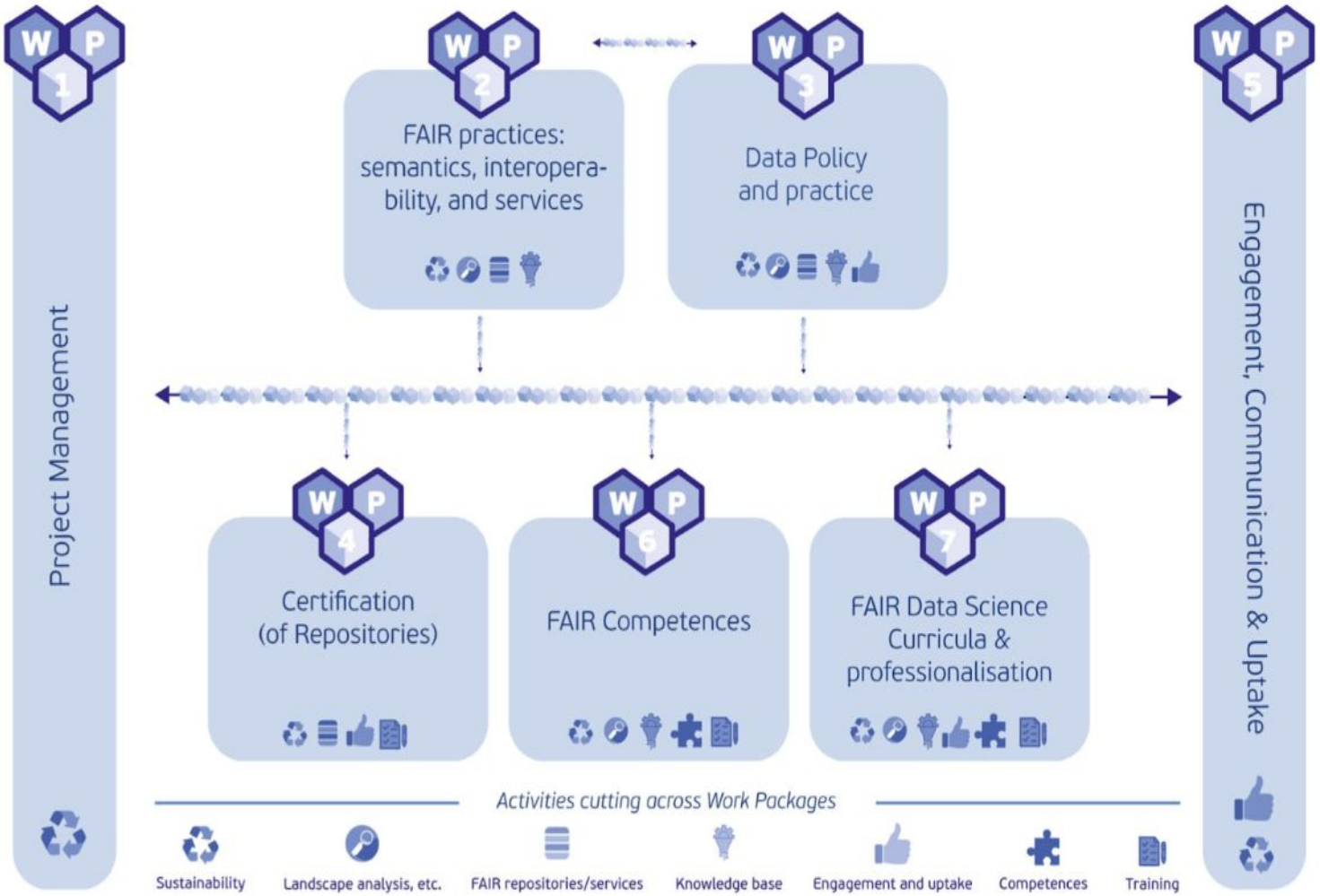
TRAINING, EDUCATION AND SUPPORT

- **Reports**
  - FAIR data in European higher education
  - Training for researchers in FAIR data science and its impact
- **FAIR competence centres tailored to different communities**
  - Three annual schools in core data skills for researchers
  - Five instructor training (train-the-trainer) events
- **FAIR competence framework for higher education**
  - Three annual FAIR data education stakeholder workshops
- **FAIR competences adoption handbook for universities**
  - Three workshops on integrating FAIR data competences
  - Case studies on good practices in FAIR competences education

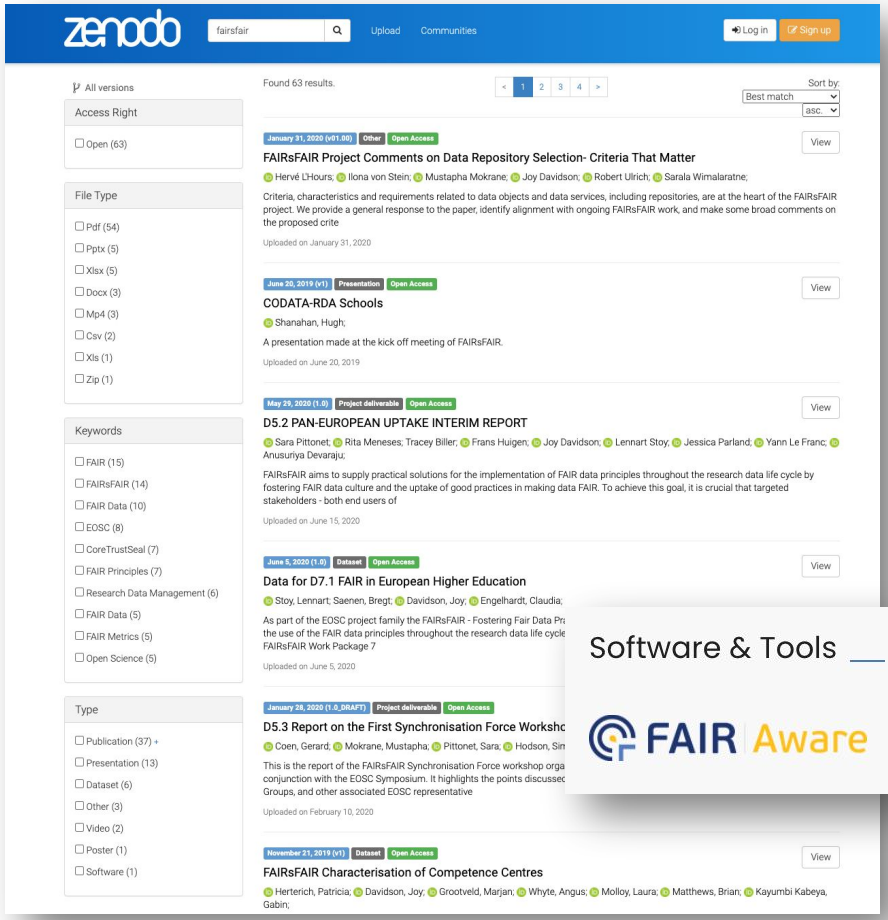
FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant Agreement 831558

FAIRSFAIR
Fostering Fair Data Practices in Europe

Work Packages



FAIRSFAIR
Fostering Fair Data Practices in Europe



Software & Tools

FAIR Aware

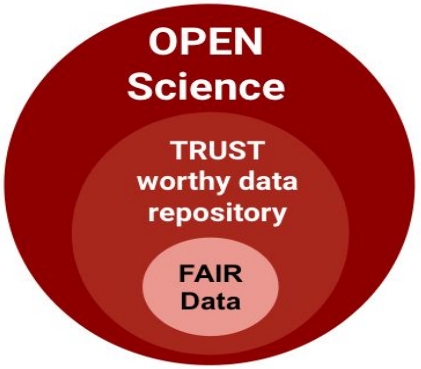
FAIR Data Object Assessment Metrics

F-UJI Automated FAIR Data Assessment Tool

Stay tuned and register: https://www.fairsfair.eu/user/register
FAIRsFAIR output on Zenodo: https://www.zenodo.org/communities/fairsfair

- We need to share our data in order to turn open science into a reality;
- The **FAIR** principles help us to define high quality and transparent research data management practices;
- Certification mechanisms, like CoreTrustSeal for digital repositories, help us to create **TRUST** in the research data infrastructure we need in order to safeguard the accessibility and assessibility of our (FAIR) data for the future.

FAIR and TRUST a perfect couple



FAIR certification

- Support the **FAIR-alignment of certification schemes**
- In-depth FAIR-enabling **repository certification support programme**
- FAIR assessment of **digital data objects**
  - F-UJI: programmatic assessment of data FAIRness
  - FAIR-Aware: tool to assess your knowledge and awareness of FAIR
- Improved registry for **finding and selecting** relevant trustworthy **repositories**

Open Call for Data Repositories

- August 2019: almost 80 submissions for the open call

- Open call: repositories to participate in the certification support programme

More information on the open calls can be found here

FAIRSFAIR
Fostering Fair Data Practices in Europe

FAIRsFAIR Certification Support Programme

Apollo - United Kingdom
DaSCH - Data and Service Center for the Humanities (DaSCH) - Switzerland
DASS-BiH (Data Archive for Social Sciences & The Humanities in Bosnia & Herzegovina) - Bosnia & Herzegovina
DASSH - The Archive for Marine Species and Habitats Data - United Kingdom
ESRF Data Repository - France
IAGOS Data Center - France
ICOS Data Portal - Sweden
The Movebank Data Repository - Germany
SOCIB - Balearic Islands Coastal Ocean Observing and Forecasting System Data Repository - Spain
Tárki Data Archive - Hungary

Certification support?

UK • DATA ARCHIVE

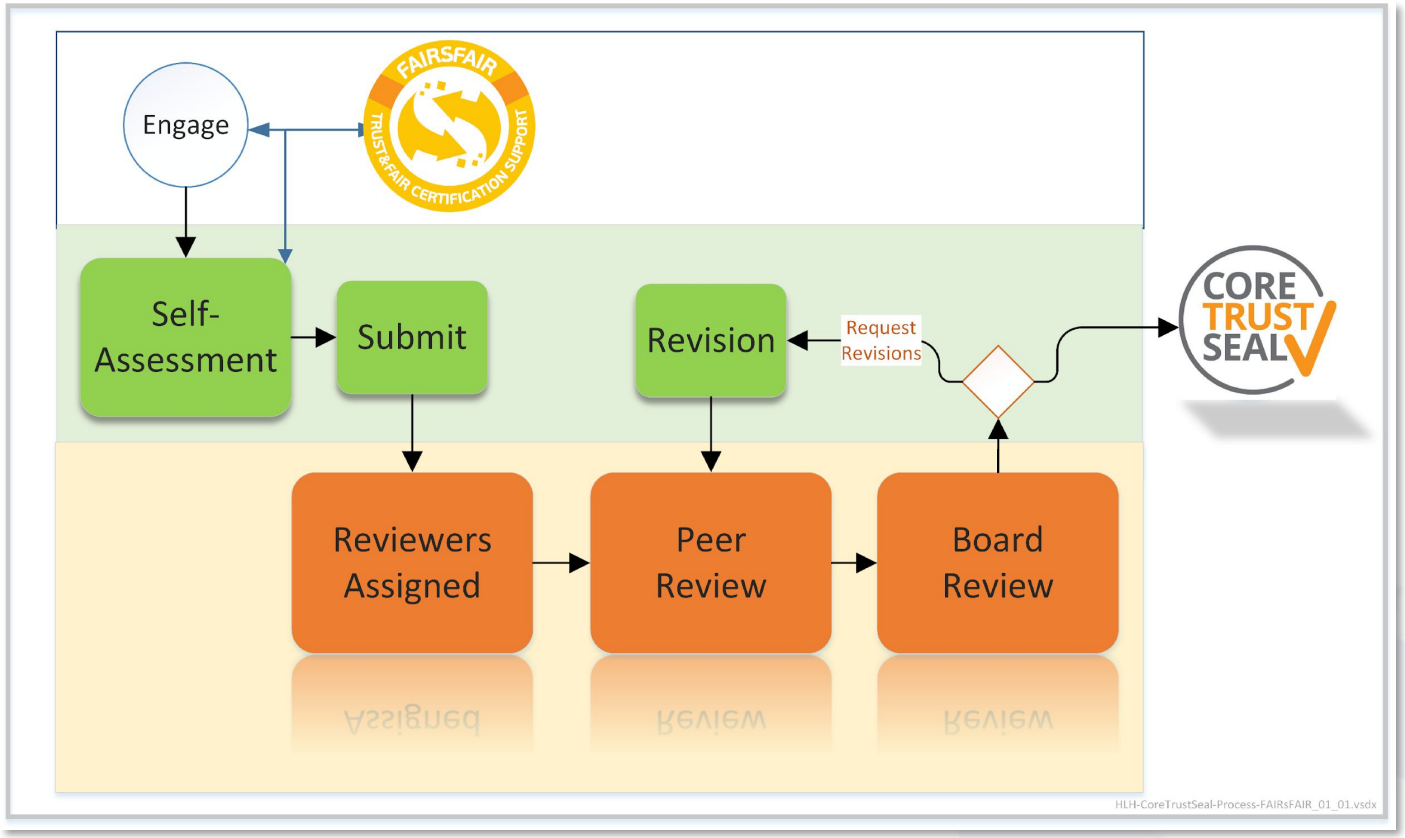
Universität Bremen

DCC

INES
Centre Informatique National de l'Enseignement Supérieur

Data Archiving and Networked Services
DANS

Certification support!

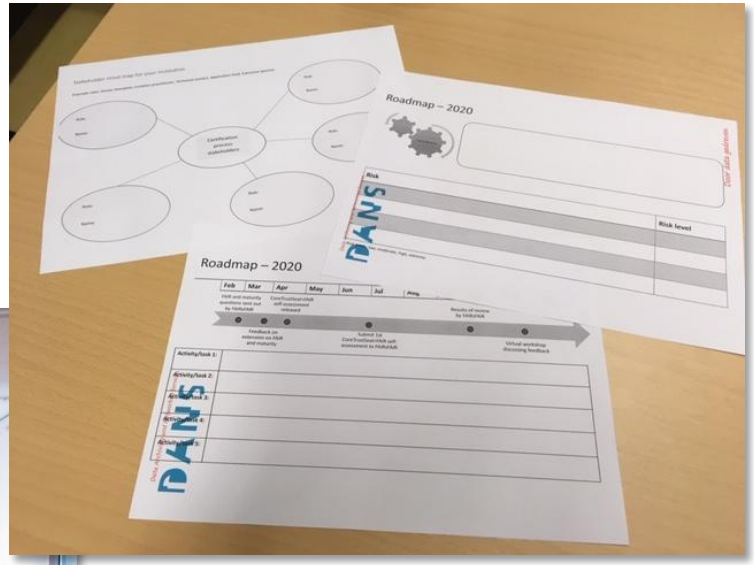


Phase 1

September 2019 - April 2020

During the Introductory workshop (February 2020) participants received foundation training in preparing for CoreTrustSeal self-assessment and were invited to comment on the CoreTrustSeal+FAIR overview to further improve the mapping between FAIR and CoreTrustSeal requirements.

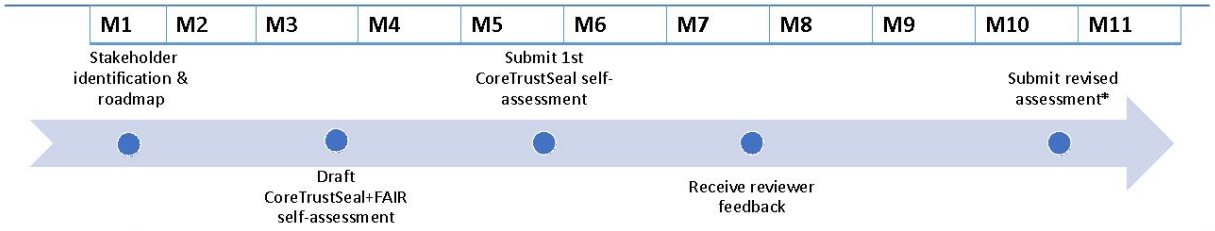
Certification Workshop



We have published the workshop presentation on Zenodo: https://doi.org/10.5281/zenodo.3754292

Road mapping exercise

Roadmap



Activity/task 1:	
Activity/task 2:	
Activity/task 3:	
Activity/task 4:	
Activity/task 5:	

* Only required if first assessment submission did not lead to successful certification

FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 proje

Roadmap

Risk	Risk level

Risk levels: low, moderate, high, extreme

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

This workshop lesson material can be found on Zenodo: https://doi.org/10.5281/zenodo.3741693

Phase 2

May 2020 - September 2020

Repositories worked on completing a draft self assessment where they answered the CoreTrustSeal requirement questions using a prepared template.

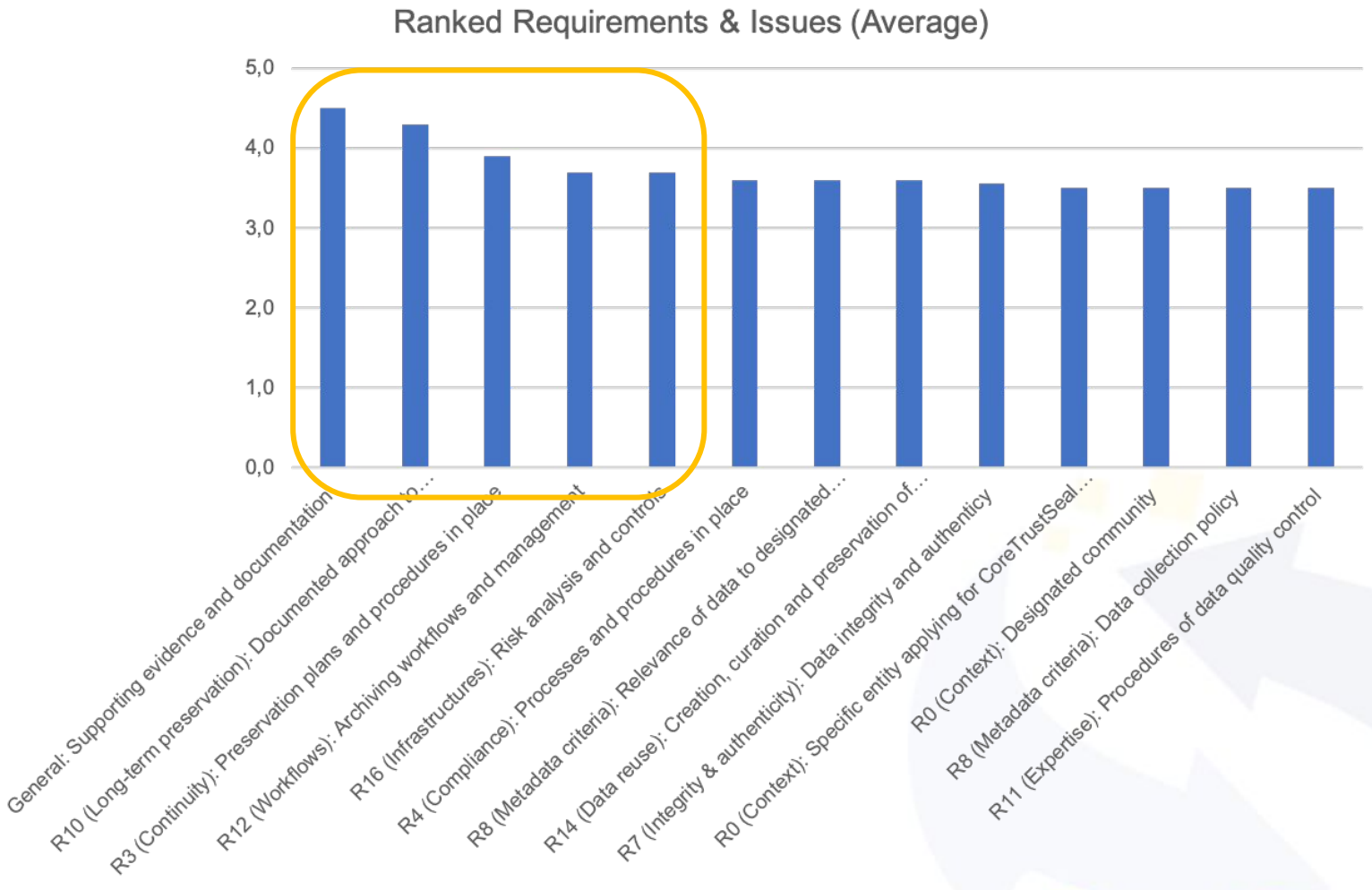
On-demand support

Phase 3

October 2020 - February 2021

The self assessments were reviewed by five teams of two reviewers each. During August and September 2020, the repositories also performed test peer reviews of their submitted work. The peer reviews will be followed by a 1-on-1-support call with the reviewer team and a further online workshop in which the general points for improvement have been addressed

Support workshop



Phase 4

February 2021 - May 2021

In February 2021, the repositories will submit their final self-assessments to the CoreTrustSeal website.

Lessons learned...

- No one-size-fits all support mode
- Expert adaptability: tailored to the repository need
- Follow-up workshops on common issues in the CoreTrustSeal self-assessment process

… and what the future brings

- Expected submissions of the self-assessments to CoreTrustSeal: January - March 2021
- Possible additional support based on the reviewers' feedback

Building capacity in the community

- FAIR-enabling repository support programme
- Data stewardship training
- FAIR training materials

FAIRSFAIR
Fostering Fair Data Practices in Europe

- Support to help repositories prepare for certification
- Support for repositories wanting to become more FAIR-enabling

Improve repository findability

- What you can do
  - Make your repository findable through re3data.org
- FAIRsFAIR support
  - Developing a registry for FAIR repositories based on DataCite's Repository finder tool
    - Filters re3data by
      - Open access data
      - Persistent Identifiers
      - Certified repository

Improve repository interoperability

- What you can do
  - Familiarise yourself with FAIR data repositories features
    - Set of FAIR data repositories features https://doi.org/10.5281/zenodo.3631528

- FAIRsFAIR support
  - Working with 12 repositories on designing implementation features to increase repository interoperability

More on how you can improve the interoperability of your repository from Claudia in a few minutes!

https://fairsfair.eu/events/prototype-implementation-fair-related-features-demonstration

Enhance FAIRness through metadata

- What you can do
  - Describe research outputs in your repository using agreed vocabularies, terminologies and metadata standards
  - Assist your depositing users in finding the most commonly used domain specific terminologies and standards

- FAIRsFAIR support
  - Providing examples of good practice in building community consensus around interoperability frameworks
  - Developing further guidance materials

https://fairsfair.eu/news/integrating-metadata-catalogues-support-cross-disciplinary-fair-uptake

Enhance FAIRness and visibility through Persistent Identifiers

- What you can do
  - Implement PIDs for different entities such as scholarly outputs, researchers, institutions, and research funders
  - Familiarise yourself with PID graphs as for example provided by the FREYA project https://www.project-freya.eu/en/pid-graph/the-pid-graph

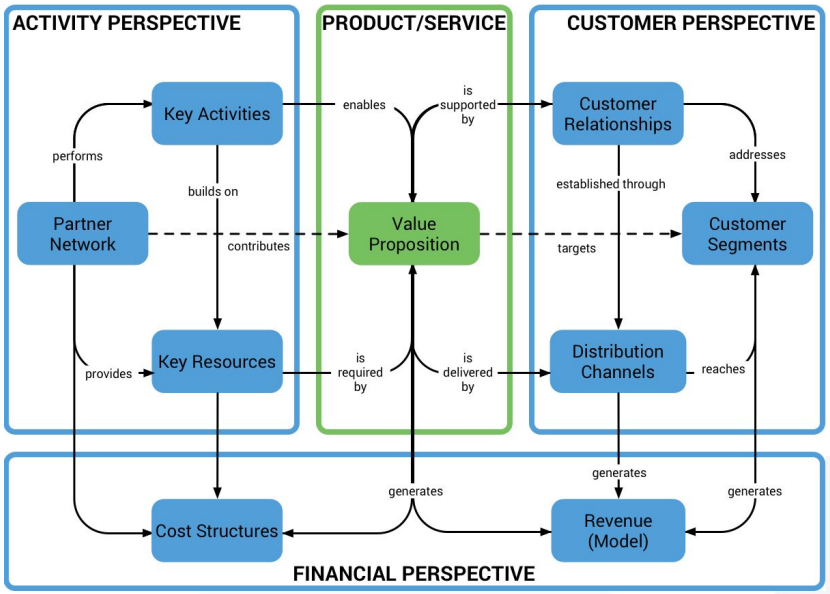
- FAIRsFAIR support
  - Working with research infrastructures to build a culture of data citation by using PIDs in their scientific activities

https://graph.openaire.eu/

Shed light on the cost of FAIR-enabling repositories

- What you can do
  - Develop an understanding of costs of your service such as costs for curating a dataset or implementing a new feature or standard supporting FAIRness
  - Publish costs e.g. for preserving data that can be included in grant applications

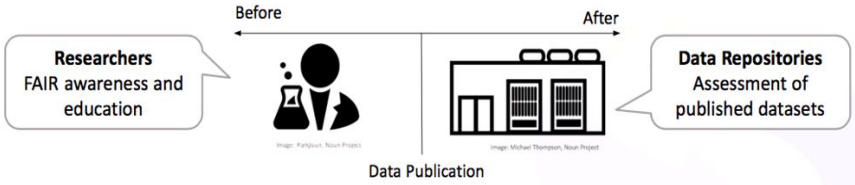
- FAIRsFAIR support
  - Developing guidance to support organisations in costing FAIR data management

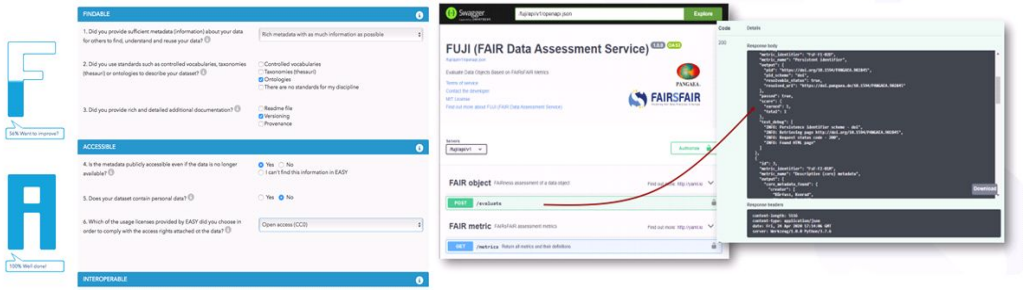


Business Model Canvas

Raise awareness of FAIR and assess the FAIRness of data

- What you can do
  - Raise awareness among researchers and data stewards (FAIR-Aware)
  - Assess the FAIRness of the data you hold (F-UJI)





- FAIRsFAIR support
  - Collecting use cases for the assessment of FAIRness of datasets
  - Developing metrics to assess the FAIRness of datasets
  - Implementing the metrics in tools addressing some of the use cases

More on these tools a bit later!

Invest in professional staff development

- What you can do
  - Explore the FAIRsFAIR training materials
  - Attend or host training on FAIR
  - Make your training and learning material FAIR:

    https://doi.org/10.1371/journal.pcbi.1007854

More on the Data Stewardship training and FAIR Data Forum from Elizabeth next.

- FAIRsFAIR support
  - Running CODATA/RDA schools

    https://fairsfair.eu/events/training
  - Developing a FAIR Competence Centre

Data Stewardship Training in FAIRsFAIR

Need **people** with the relevant skills to help make your repository more FAIR-enabling.

Researchers need skills to practice FAIR

Need training on data stewardship and specialist training for “data stewards”

Need access to training materials

Building on the CODATA-RDA School - developed new data stewards strand

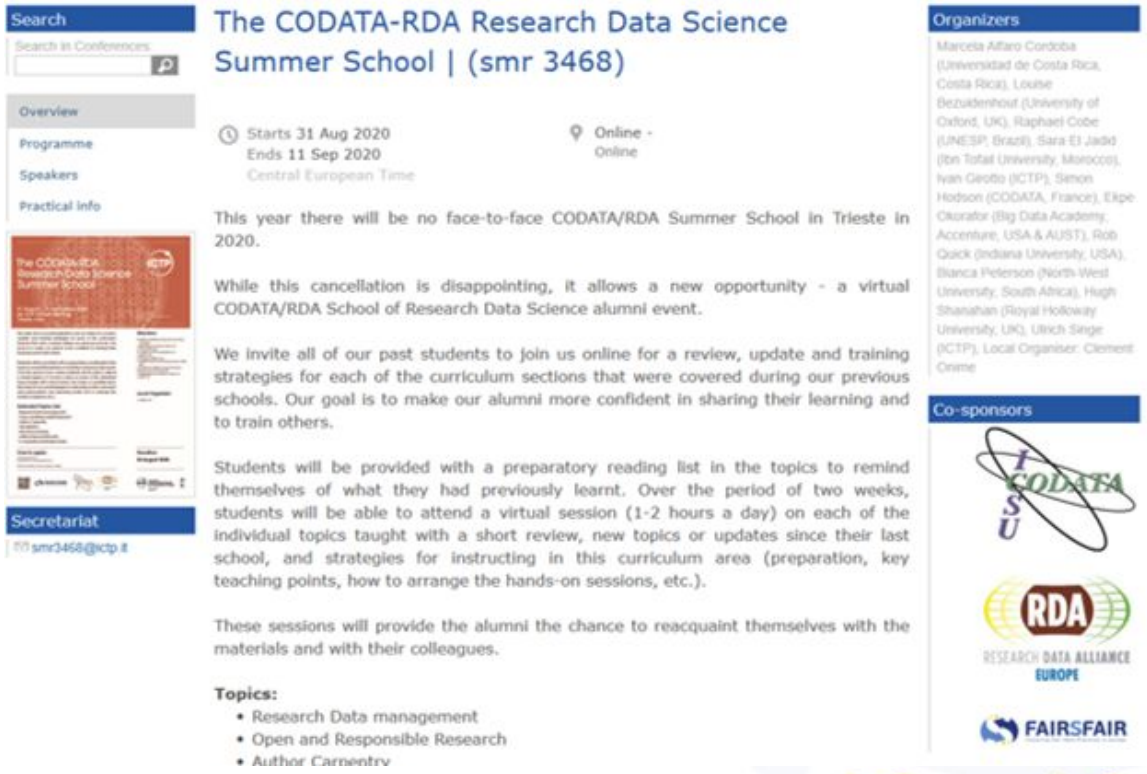
First school ran August 2019

Two weeks residential school

First week common curriculum for Early Career Researchers and Data Stewards, second week split with separate programme for Data Stewards

Further information on this: http://doi.org/10.5281/zenodo.3609205

- Next school was planned for August 2020 but due to Covid a rethink was necessary
- Virtual Trieste online event for Alumni of previous CODATA-RDA Data Science Schools, included sessions on RDM and Open and Responsible Research



Search

Search in Conferences

Overview

Programme

Speakers

Practical Info

Secretariat

smr3468@ictp.it

The CODATA-RDA Research Data Science Summer School | (smr 3468)

Starts 31 Aug 2020
Ends 11 Sep 2020
Central European Time

Online -
Online

This year there will be no face-to-face CODATA/RDA Summer School in Trieste in 2020.

While this cancellation is disappointing, it allows a new opportunity - a virtual CODATA/RDA School of Research Data Science alumni event.

We invite all of our past students to join us online for a review, update and training strategies for each of the curriculum sections that were covered during our previous schools. Our goal is to make our alumni more confident in sharing their learning and to train others.

Students will be provided with a preparatory reading list in the topics to remind themselves of what they had previously learnt. Over the period of two weeks, students will be able to attend a virtual session (1-2 hours a day) on each of the individual topics taught with a short review, new topics or updates since their last school, and strategies for instructing in this curriculum area (preparation, key teaching points, how to arrange the hands-on sessions, etc.).

These sessions will provide the alumni the chance to reacquaint themselves with the materials and with their colleagues.

Topics:

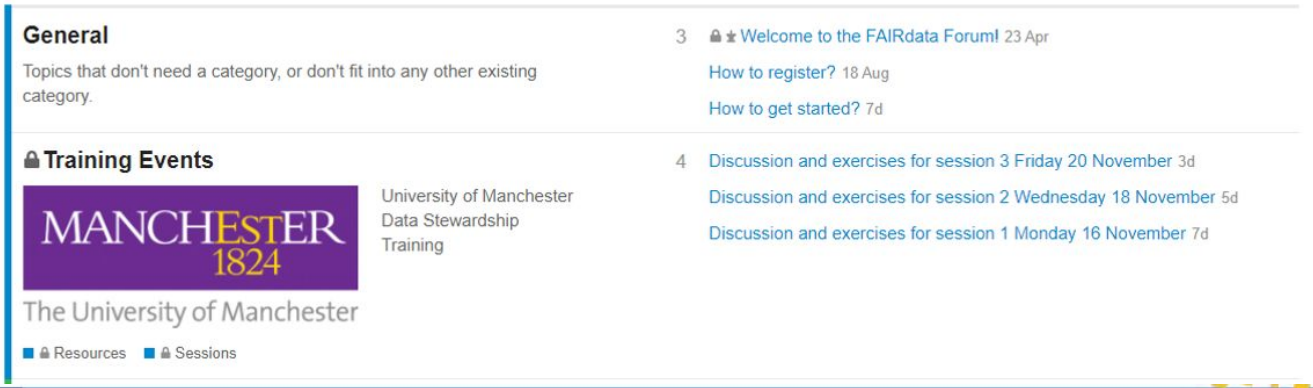
- Research Data management
- Open and Responsible Research
- Author Carpentry

Organizers

Co-sponsors

Developing the training model for HEIs

- University of Manchester, November 2020
  - 14 research data management professionals
  - worked with the local organisers and FAIRsFAIR instructors to agree a programme from the core list of topics
  - 3 hours of training over the course of a week
  - flipped learning approach - materials made available in advance of the sessions
  - opportunity for discussion outside of the training session via shared space on the FAIRdata Forum (https://fairdataforum.org)



What topics are covered?

Core list of topics that can be adapted for different training scenarios/groups:

- Research Data Management, Open Science, and FAIR Principles
- What is Data Stewardship?
- Development of Research Data Management Policies
- Promoting and Archiving Data
- Ontologies
- Open and Responsible Research

Programme is evolving. Need to adapt materials for different levels e.g. beginner/intermediate/advanced

FAIR Training Materials - working with the community

Alongside training as a community we need access to training materials that can be re-used.

How can we develop FAIR Training materials?

- RDA Education and Training on handing of research data - Interest Group https://www.rd-alliance.org/groups/education-and-training-handling-research-data.html
  - making training materials findable - minimal/extended metadata for training materials
- terms4FAIRskills https://terms4fairskills.github.io/
  - providing a use case of training materials to help develop the terminology

https://github.com/FAIRDataTeam/FAIRDataPoint-Spec

FAIRSFAIR

Fostering Fair Data Practices in Europe

Implementing a FAIR Data Point - lessons learned

Making your repository more FAIR-enabling

Claudia Behnke (SURF)

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

FAIRSFAIR

Fostering Fair Data Practices in Europe

Improving a FAIR Data Point - still learning

Making your repository more FAIR-enabling

Claudia Behnke (SURF)

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

Why?

We asked 100* people…

D2.1 Report on FAIR requirements for persistence and interoperability 2019
D2.4 2nd Report on FAIR requirements for persistence and interoperability

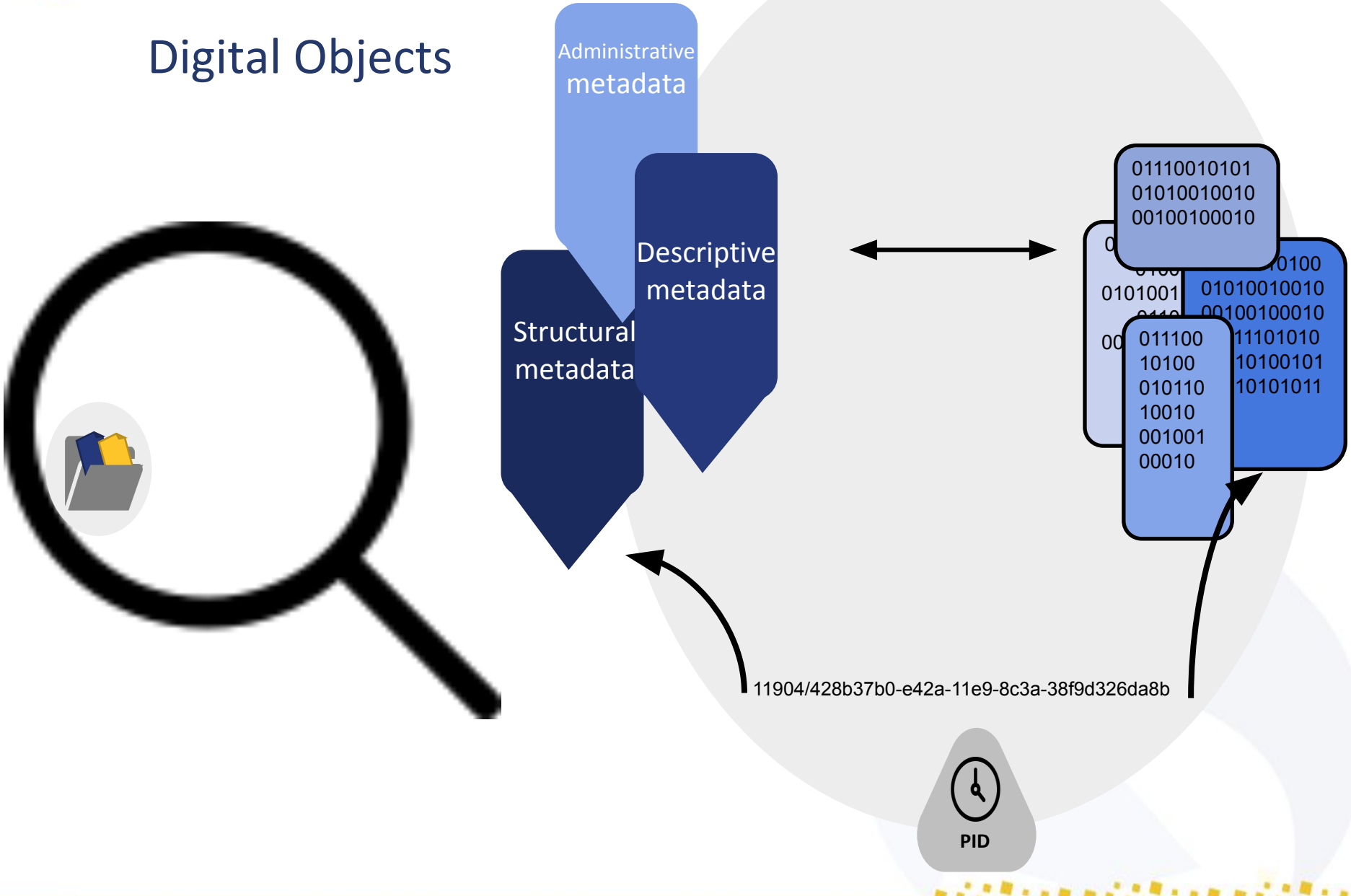
D2.2 FAIR Semantics: First recommendations

M2.15 Assessment report on 'FAIRness of software'

D2.3 Set of FAIR data repositories features

Wittenburg et al., 2018

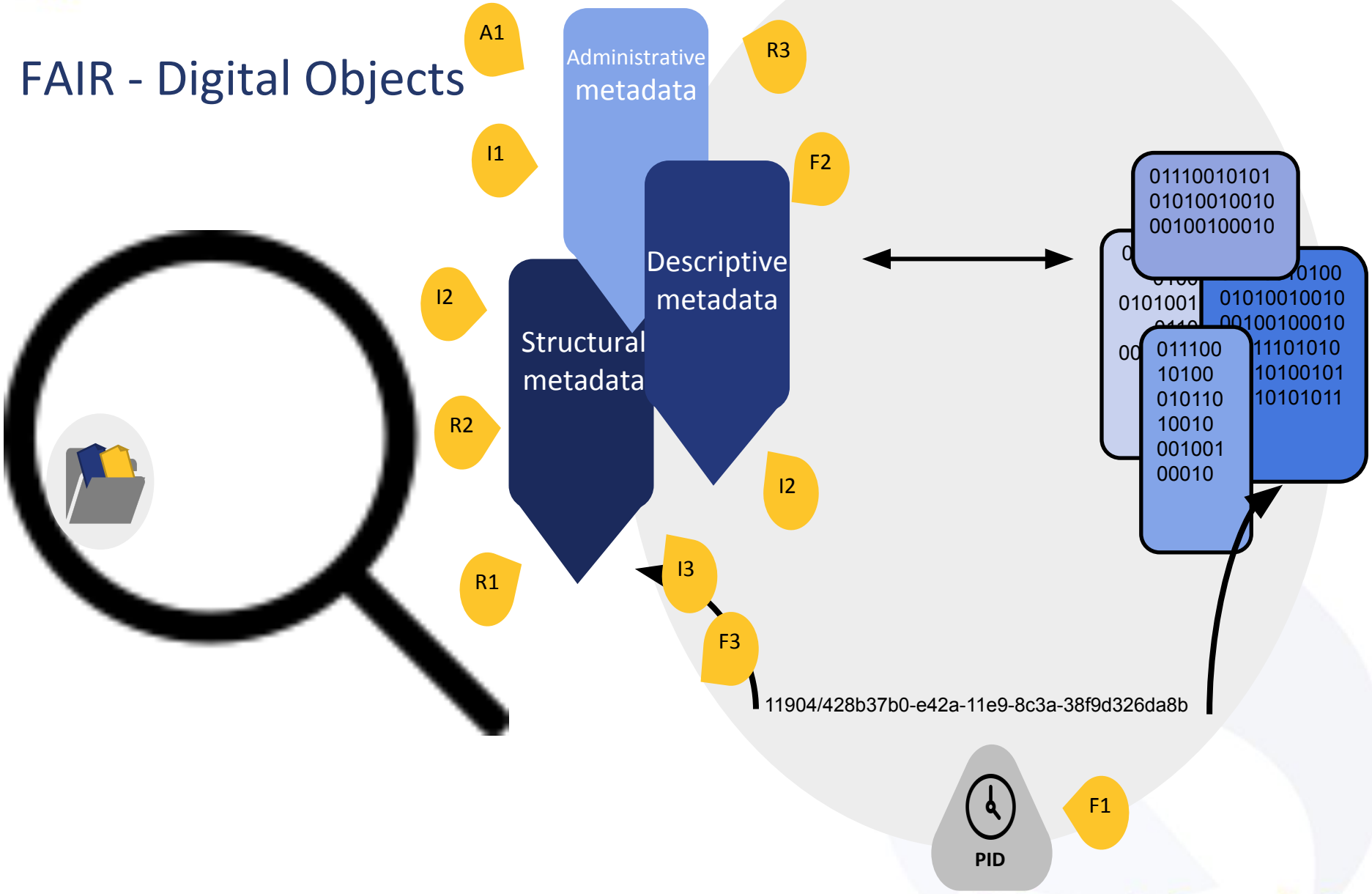
Digital Objects



Wittenburg et al., 2018

Wilkinson et al., 2016

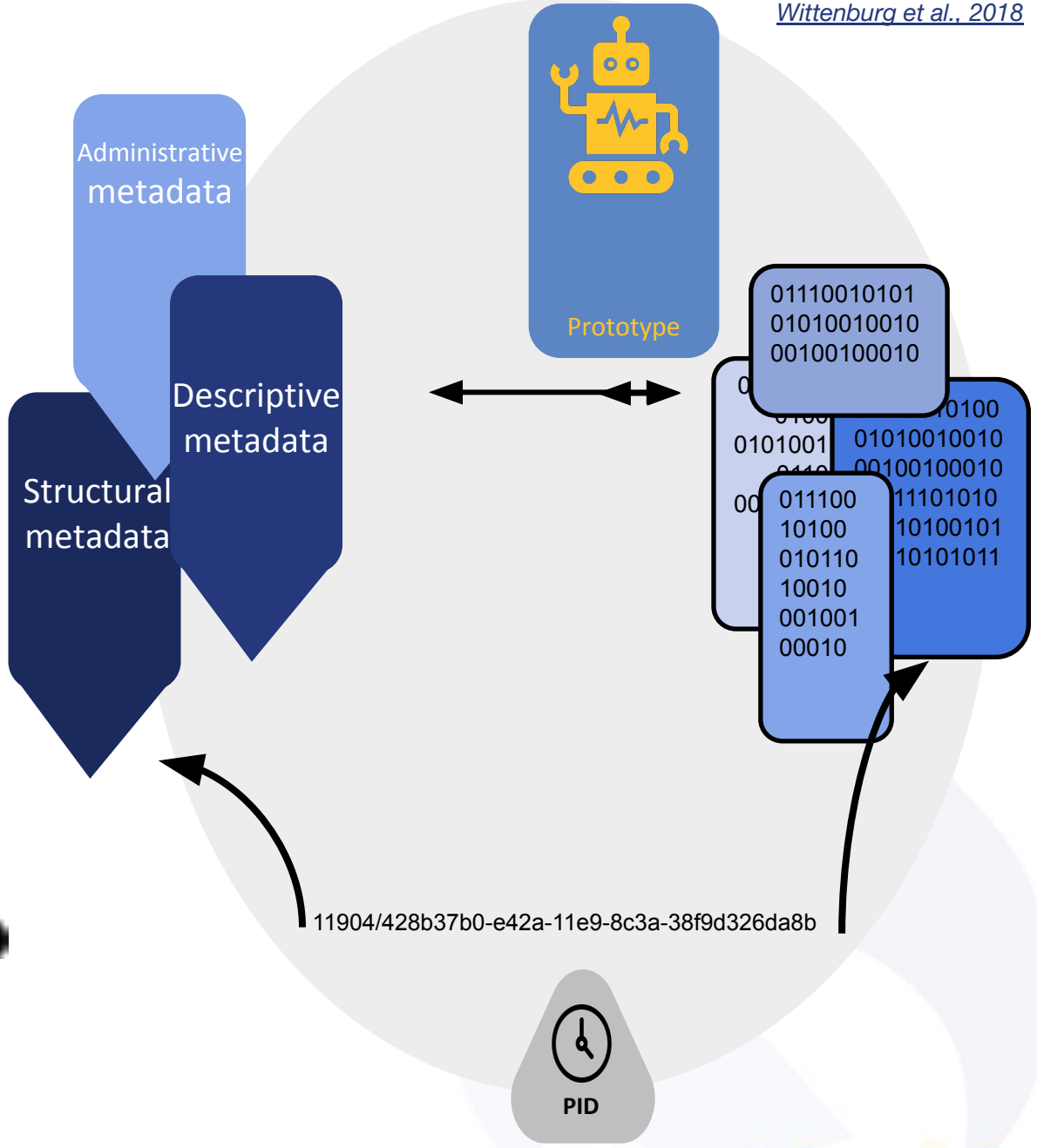
FAIR - Digital Objects



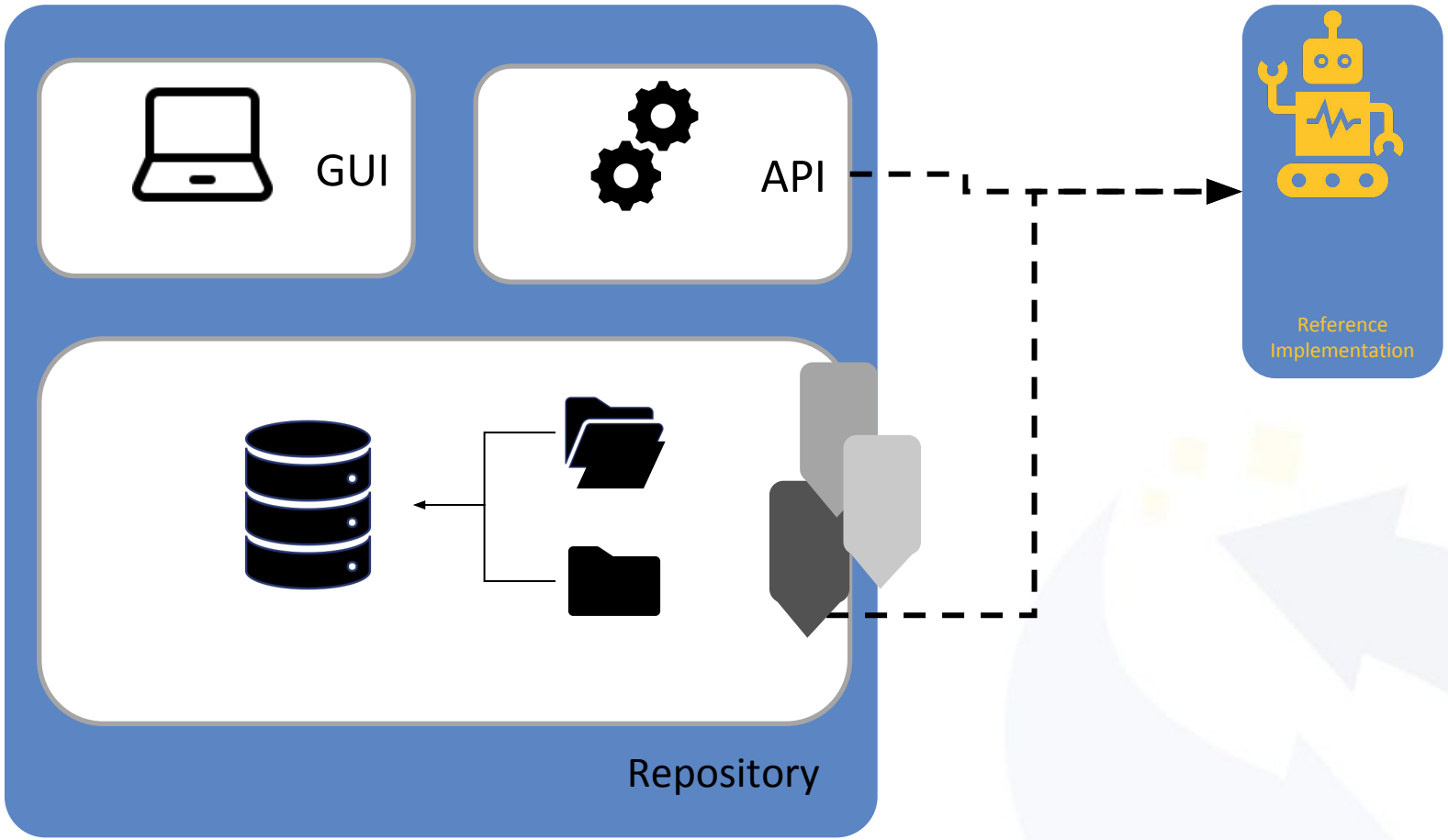
Lesson 1:
It's all about metadata, stupid!

Wittenburg et al., 2018

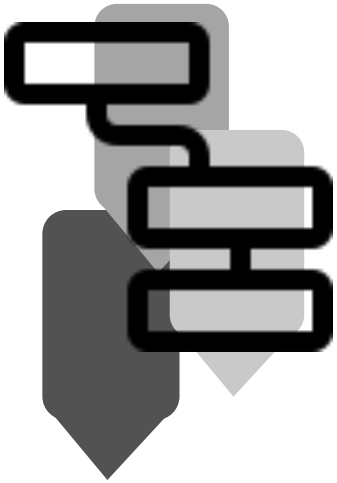
FAIR - Digital Objects

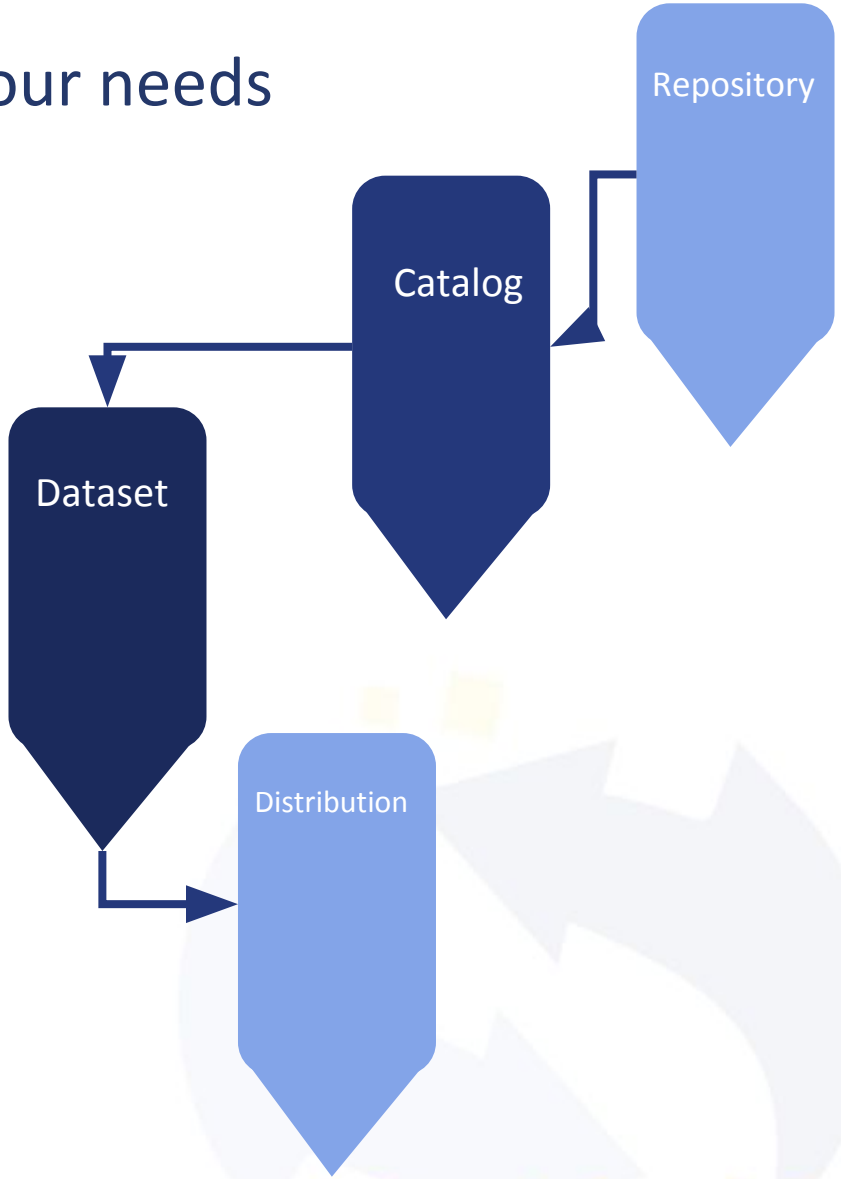


How would this help repositories & users



DCAT2: extend metadata to your needs

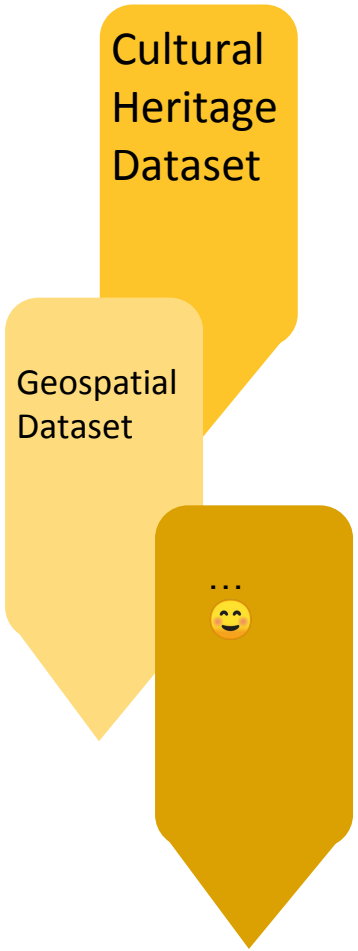


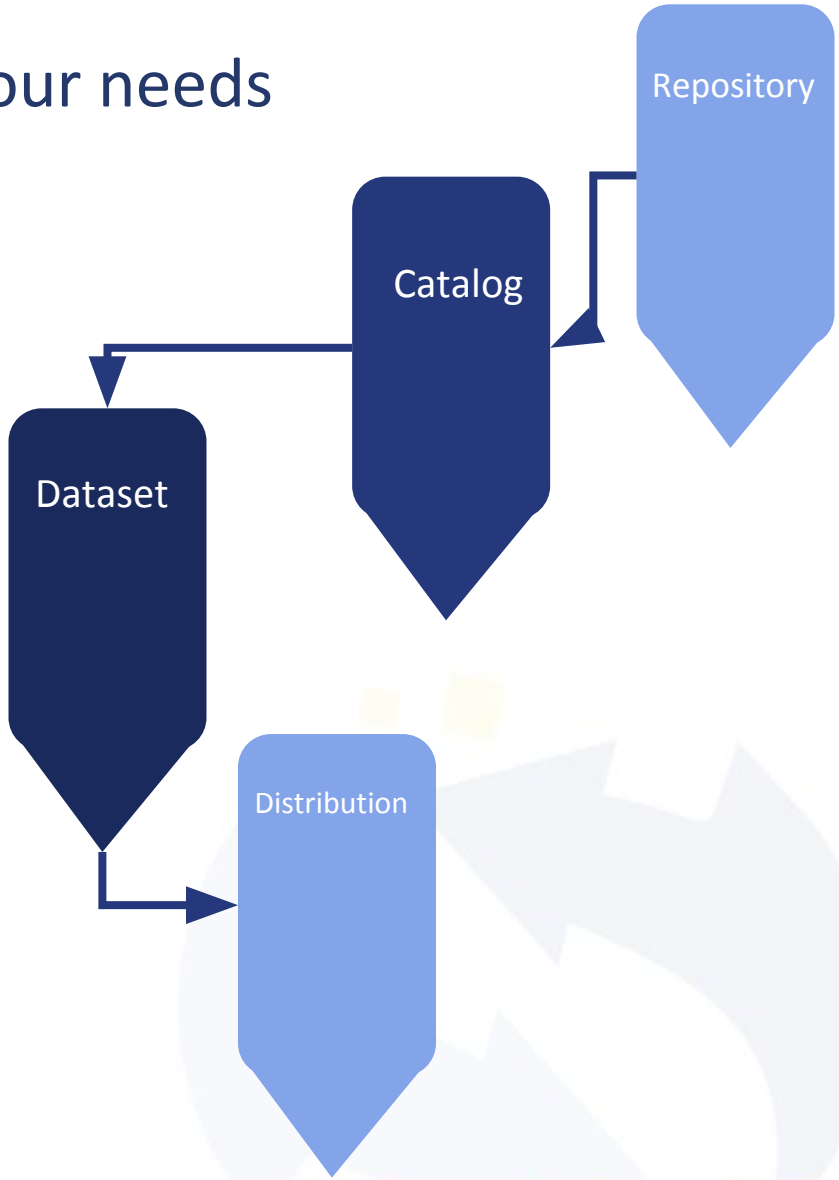


Lesson 2: There is no common metadata standard

https://www.w3.org/TR/vocab-dcat-2/#vocabulary-overview

DCAT2: extend metadata to your needs





Lesson 3: Repositories know best what they need

FAIRSFAIR
Fostering Fair Data Practices in Europe

Prototype implementation of the FAIR-related features - a demonstration

Enhancing Metadata Exposure in Repositories

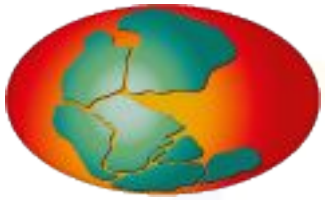
Online workshop

9 December 2020

15:00-18:00 CET

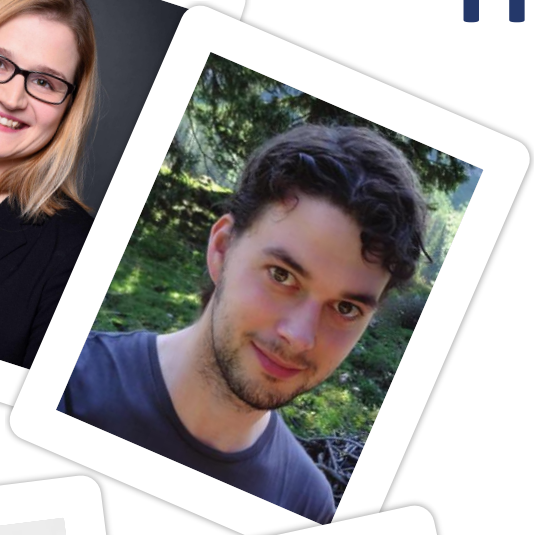
REPOSITORIES

- "Developer"
- "Tester"

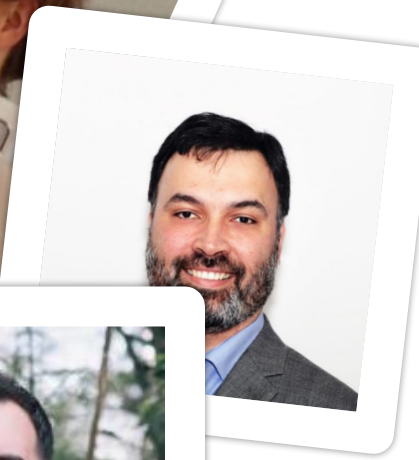
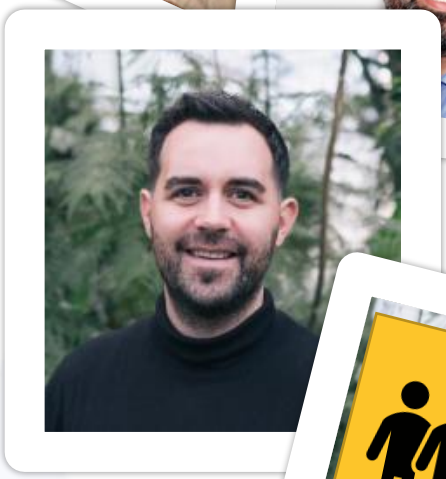


FAIRSFAIR
Fostering Fair Data Practices in Europe

THE TEAM




CSC

Data Archiving and Networked Services
DANS

DTL

esciencedata FACTORY

GO FAIR

SURF SARA

FAIRSFAIR
Fostering Fair Data Practices in Europe

F-UJI : An Automated Assessment Tool for Improving the FAIRness of Research Data

Robert Huber & Anusuriya Devaraju 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

Task 4.5 – FAIR Data Assessments: 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

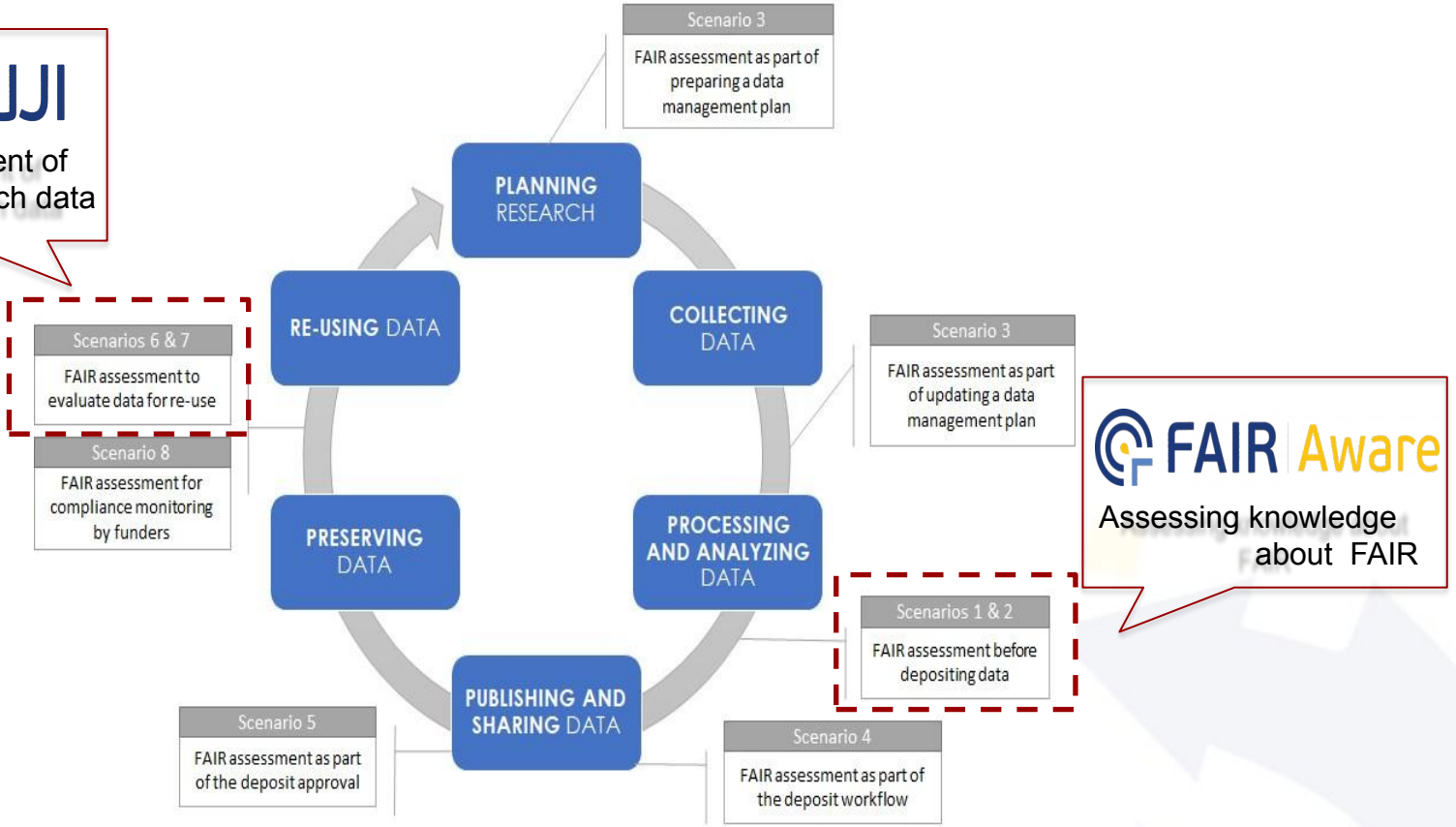
Supporting Recommendations
Rec. 25: Implement FAIR metrics to monitor uptake

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

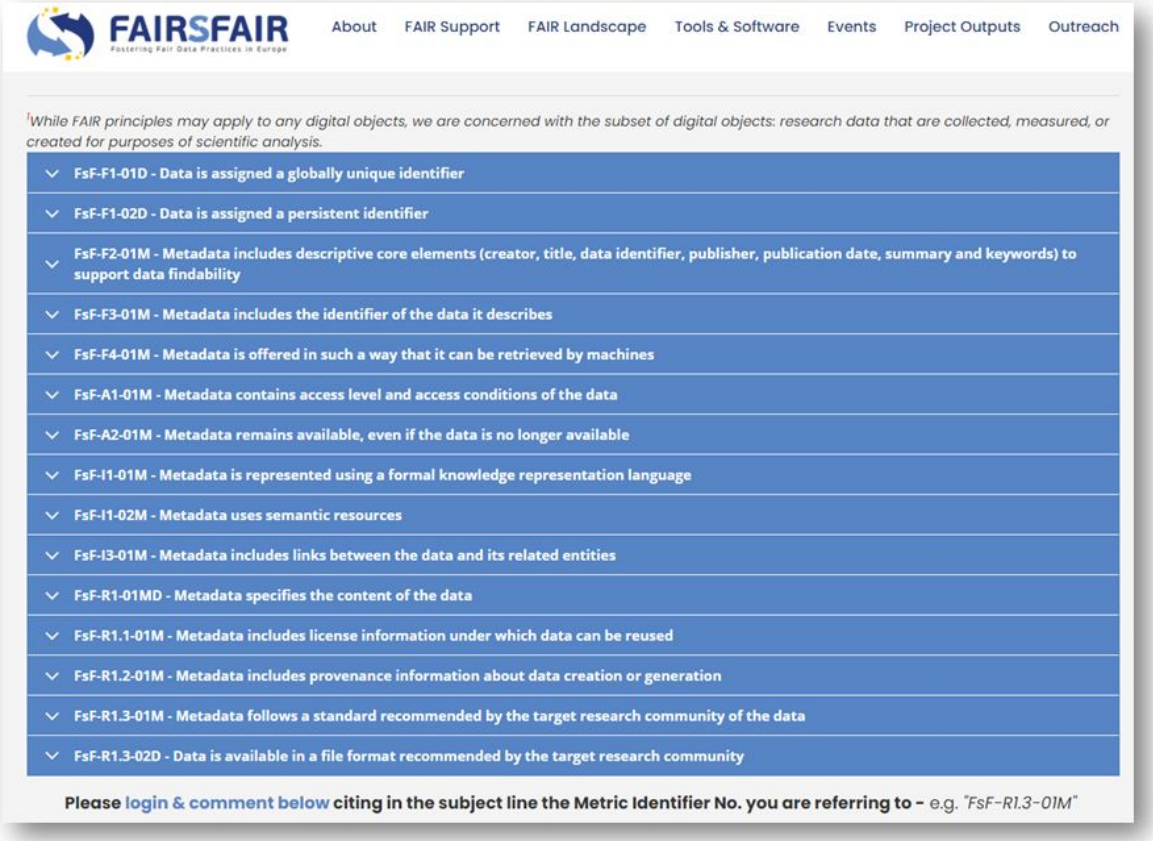
F-UJI
FAIR assessment of published research data



FAIR Aware
Assessing knowledge about FAIR

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

FAIR assessment metrics

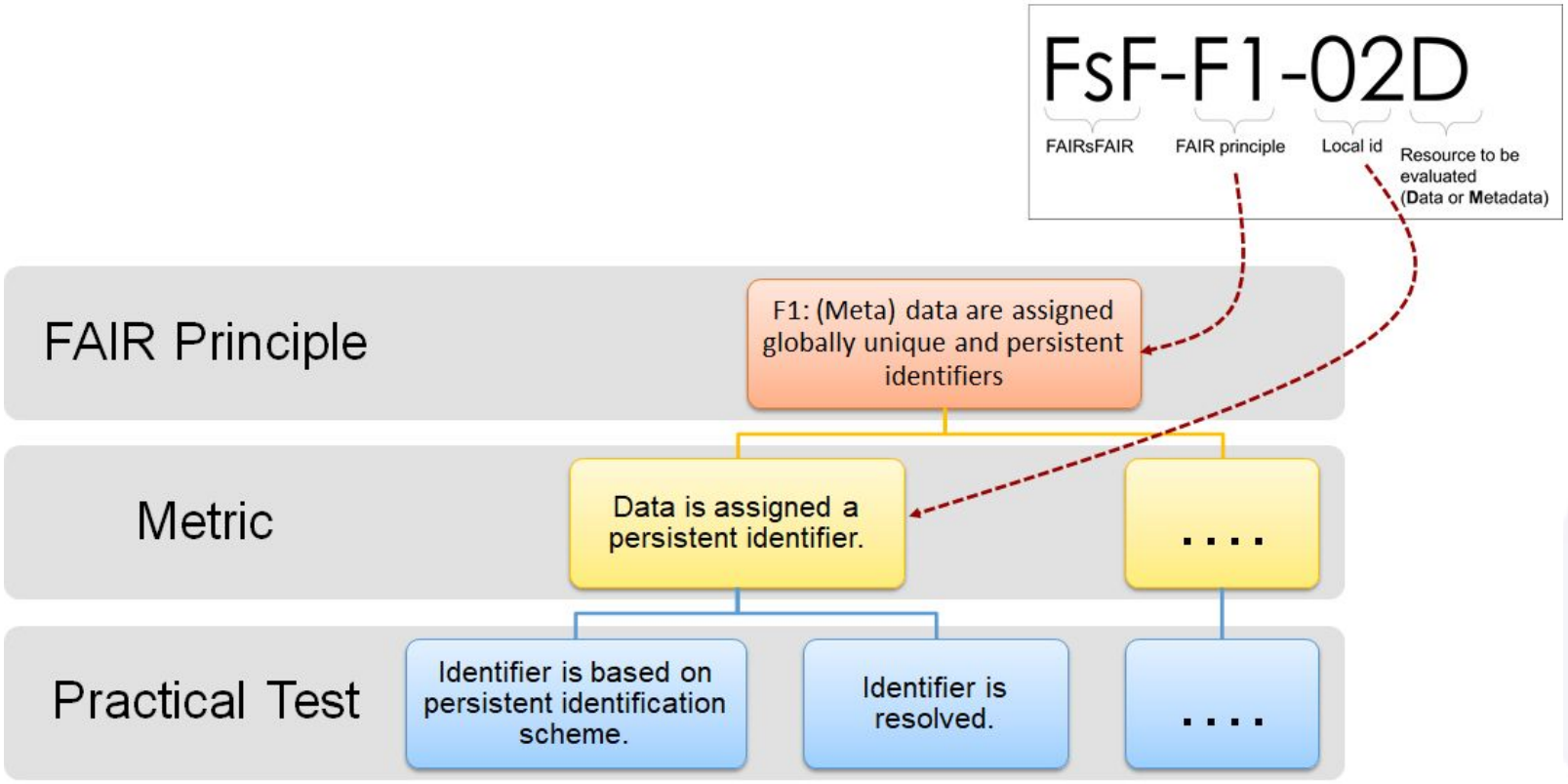


Object Assessment Metrics v0.4

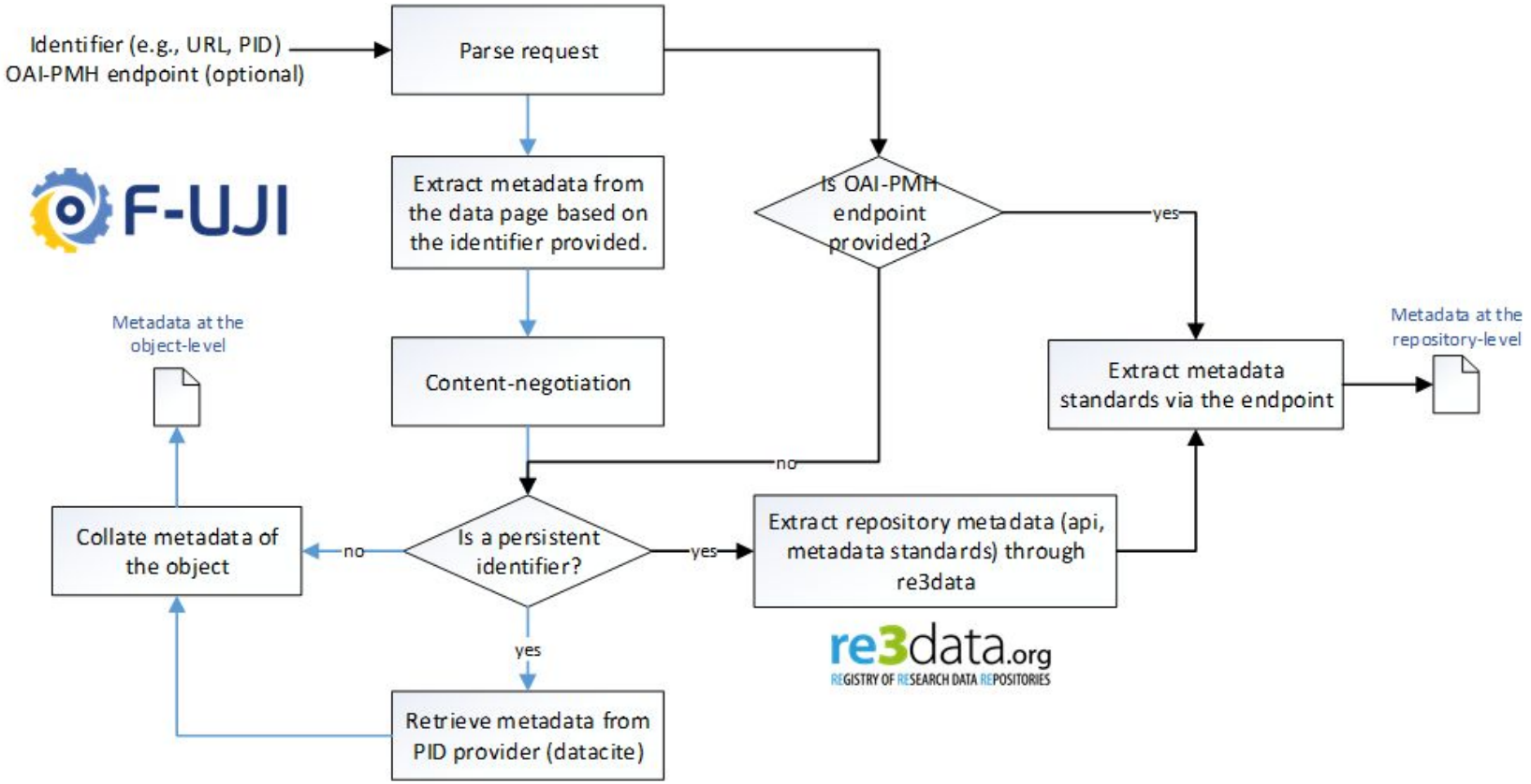
We would love to hear your feedback!

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

From Principles to Practical Tests



High Level Flow (Data Gathering)



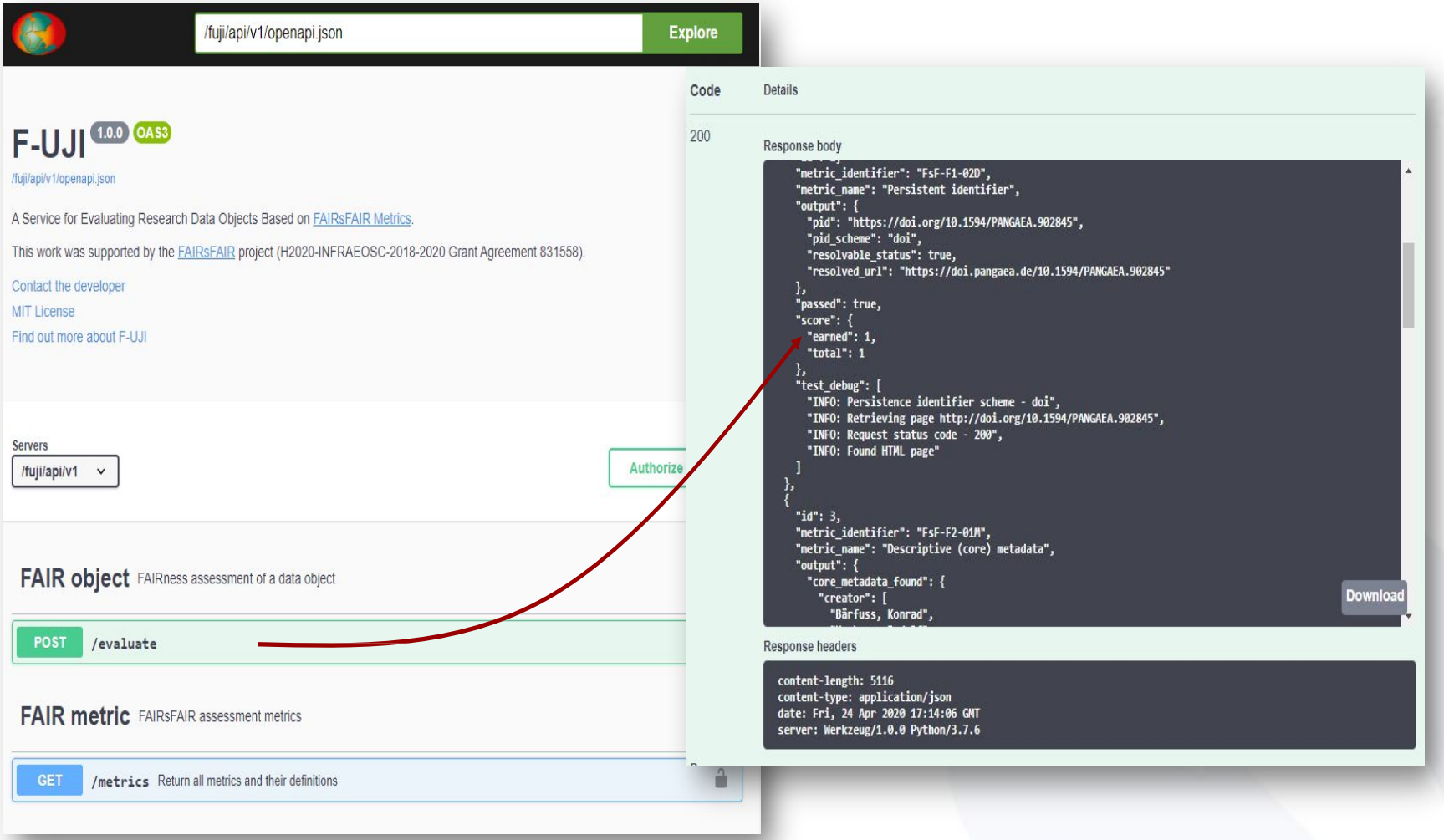
FAIRSFAIR
Fostering Fair Data Practices in Europe

F-UJI

DEMO

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

F-UJI – An Automated FAIR Data Assessment Tool



F-UJI – An Automated FAIR Data Assessment Tool

F-UJI Automated FAIR Data Assessment Tool

URL/PID: https://doi.org/10.17863/CAM.14473

OAI-PMH: (Optional) Enter the URL of an OAI-PMH endpoint ser

run test

Use DataCite metadata? Caching results?

Assessment Results:

Evaluated Ressource:

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).

Title:	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).
Ressource ID:	https://doi.org/10.17863/CAM.14473

Summary:

Findable: 4 of 7

Accessible: 1 of 3

Interoperable: 0 of 4

Reusable: 2 of 9

Report:

Findable

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

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

Status: pass

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"
}
```


Debug:

Level:	Message:
INFO	PID schemes-based assessment supported by the assessment service - ['doi', 'handle', 'ark', 'purl', 'lsid', 'sra', 'biosample', 'ensembl', 'uniprot', 'genome', 'urn']
INFO	Retrieving page http://doi.org/10.17863/CAM.14473
INFO	Content negotiation accept=text/html, application/xhtml+xml, 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.

Accessible

FAIR data takes work

Challenges at all levels

Findable?

Accessible?

Interoperable?

Reusable?

Part of the solution

Parts of the presentation are based on an earlier given workshop "FAIR-Aware: From depositing to discovering data". The workshop was prepared by Linas Cepinskas, Kathleen Gregory and Charlotte Glas and delivered during "De Nederlandse Dataprijs: Together we share" on 18 November 2020.

FAIR-Aware: Part of the solution

- Practical
- Accessible
- Educational
- For anyone working with datasets

https://fairaware.dans.knaw.nl

DEMO

https://fairaware.dans.knaw.nl

Making the work of FAIR data easier

- Finding, understanding and reusing data involve the work of *humans* (as well as *machines*).
- FAIR-Aware helps with human understanding & assessment with the goal of increasing the FAIRness of data for the purpose of *reuse.*

https://fairaware.dans.knaw.nl

Questions and discussion

Looking at the chat, questions arise concerning...

… metadata schemes
… certification
… institutional repositories
… (use of) FAIR assessment tools
… supported APIs
…

Raise your hand and you will be given the floor!
😊

Wrapping up

FAIRsFAIR is about collaboration: experts working on questions surrounding various aspects of FAIR data and Open Science.

On social and technical level, as well as business improvement and software.

From a user perspective, but also from the perspective of the facilitator: capacity building and expert guidance.