RDA FAIR data maturity model

16th of September 2020

Workshop #10

Agenda

- Welcome

- FAIR assessments and DMPs

- FAIR use cases

- Next steps

- Closing

Welcome

| Meeting | State of play | Roadmap (remainder 2020) |

Meeting

1. State of play

2. Explore FAIR assessments in the context of (i) DMPs and (ii) funding agencies

3. Further action

State of play

- Publication of the [FAIR Data Maturity Model: Specification and Guidelines](#) as an RDA recommendation, 25 June 2020

- EOSC-SYNERGY [Intermediate report on technical framework for FAIR principles implementation](#), 2 Sept. 2020

Roadmap (remainder 2020)

- Workshop #10
- Bridging the gap between Funders and Communities (survey)
- Workshop #11
- Establish work plan for 2021
- CODATA FAIR symposium

FAIR assessment and DMPs

Hands-on presentations on the role of DMPs and how DMP tools can help to achieve FAIRness

10' — Data Stewardship Wizard

Rob Hooft
DTLS

10' — DMPs and FAIRsFAIR

Angus Whyte
Patricia Herterich
FAIRsFAIR

10' — Open discussion
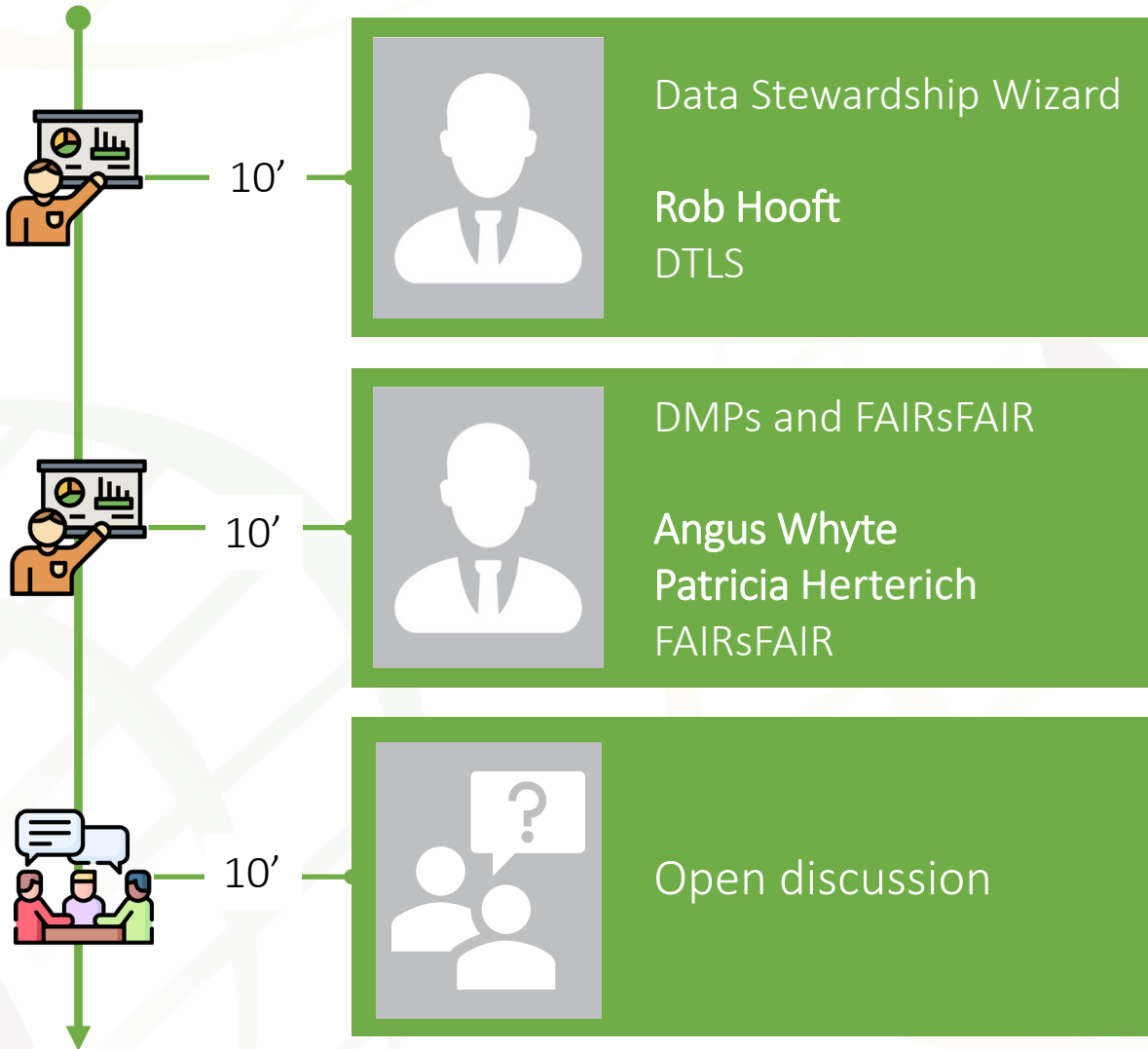
Predicting FAIRness using DS Wizard

Rob Hooft, Dutch Techcentre for Life Sciences

FAIR Data Maturity model WG, 2020-09-16

When can we measure FAIRness?

- The FAIR Maturity model asks questions in the present tense

- Existing automated methods (objective analysis) for FAIR evaluation test an existing document or URL

- What if the data does not exist yet?

What if the data does not exist yet?

- Data stewards need to help researchers achieve FAIRness
  - Provide guidance
  - Help balance investment and benefits
  - Predict how FAIR data will become

- A good place to do that is in the Data Management Plan
  - Start while writing the project proposal
  - Details at the beginning of the project execution
  - Fill in results at the end of the project

DSW

Data Stewardship Wizard

Create Smart Data Management Plans for FAIR Open Science

🧙 Get started

🧙 DS Wizard   My Experiment (Common DSW Knowledge Model, 1.4.0)   Save

I. Design of experiment

⛓ Knowledge Models

Current Phase

📋 Questionnaires   Before Submitting the Proposal ▾

Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others.

☑ KM Editor

DS Wizard? Brief feature-overview

- Expert system for DMP creation: Query engine + Knowledge Model
  - For data stewards: an editor for knowledge models
- Trying to turn DMP creation from Burden to Benefit for a project
- An Expert does not ask the same questions to everyone
  - DS Wizard has hierarchical questionnaire for context-dependent questions
  - It has many questions, but most of them are "multiple choice"
- Supports the RDA common model for machine actionable DMPs
- Supports DMP as living document
- Capable of "writing a DMP" for a funder based on the answers

For FAIR Maturity WG (Common DSW Knowledge Model, 2.2.0)

(unsaved changes)   Discard   **Save**

- Knowledge Model Editor
- Knowledge Models
- Questionnaires
- Documents
- Storage Costs Evaluator

Current Phase

Before Submitting the Proposal ⇕

Chapters

I. Administrative details	1
II. Re-using data	1
III. Creating and collecting data	**9**
IV. Processing data	3
V. Interpreting data	1
VI. Preserving data	6
VII. Giving access to data	3

More

Summary Report

Help ›

Rob Hooft ›

« Collapse sidebar

2 **Will you be using new types of data?**

Sometimes the type of data you collect can not be stored in a commonly used data format. In such cases you may need to make your own, keeping interoperability as high as possible.

☑ Desirable: *Before Submitting the Proposal*
▤ Data Stewardship for Open Science: *ikk*

○ **a. No, all of my data will fit in common formats**
  `Interoperability`

◉ **b. Yes, I will need to use custom formats for some of my data** ☰

↺ Clear answer

2.b.1 **Will you need to add fields in your data format to a data type registry?**

Even if the data format you are using is unique to your project, the discrete data items should be reused or reusable as much as possible. Data type registries can help with that.

☑ Desirable: *Before Submitting the Proposal*

○ **a. No, all of my data types are described in a data type registry already**
  `Interoperability`

DS Wizard

For FAIR Maturity WG (Common DSW Knowledge Model, 2.2.0)

(unsaved changes)   Discard   **Save**

- Knowledge Model Editor
- Knowledge Models
- Questionnaires
- Documents
- Storage Costs Evaluator

Current Phase

Before Submitting the Proposal ⬍

Chapters

I. Administrative details	1
II. Re-using data	1
III. Creating and collecting data	9
IV. Processing data	3
V. Interpreting data	1
VI. Preserving data	6
VII. Giving access to data	3

More

Summary Report

- Help ›
- Rob Hooft ›
- « Collapse sidebar

2.b.4 How will you design your new data format?

☑ Desirable: *Before Submitting the Proposal*

▤ Data Stewardship for Open Science: *jwg*

○ **a. There is a closely related more generic and open format that I can specialize**

   Interoperability

○ **b. I will use a Linked Data format**

   Interoperability

○ **c. I will use a completely custom data format**

   Interoperability

2.b.5 Will you describe your new data format for others?

☑ Desirable: *Before Submitting the DMP*

○ **a. No, this is not needed**

   Interoperability

○ **b. Yes, I will register my standards at a data standards registry**

KNOWLEDGE MODEL   TAGS   PREVIEW

... 〉 Will you be using n... 〉 Yes, I will need to u... 〉 How will you design... 〉 There is a closely r...

DS Wizard

- Knowledge Model Editor
- Knowledge Models
- Questionnaires
- Documents
- Storage Costs Evaluator

Close

○ Number of files/subjects
○ Rough average size of each file/
▢ njy
▾ ○ Will you be using new types of data?
  ☑ No, all of my data will fit in common fo
  ▾ ☑ Yes, I will need to use custom formats
    ▾ ○ Will you need to add fields in your d
      ☑ No, all of my data types are desc
      ☑ Yes, I will add new types to an e:
      ☑ Yes, I will create my own data ty
    ○ Which data type registries will you
    ▾ ○ Do you need to create vocabularies
      ☑ No, suitable public controlled vo
      ☑ Yes, I will make and publish a vo
      ▢ ske
  ▾ ○ How will you design your new data
    ☑ There is a closely related more g
    ☑ I will use a Linked Data format
    ☑ I will use a completely custom d:
    ▢ jwg
  ▾ ○ Will you describe your new data for
    ☑ No, this is not needed
    ☑ Yes, I will register my standards
▢ ikk

You can use Markdown and see the result in the Preview tab.

Follow-up Questions

+ Add follow-up question

Metrics

( ) **Findability**

( ) **Accessibility**

(●) **Interoperability**

Weight [1]   Measure [0.5]

( ) **Reusability**

( ) **Good DMP Practice**

( ) **Openness**

Help >

Rob Hooft >

« Collapse sidebar

DS Wizard

✕ Close   ↴ Create Document   More ▾

- ✎ Knowledge Model Editor
- ⬡ Knowledge Models
- ▤ Questionnaires
- ▤ Documents
- ▭ Storage Costs Evaluator

Current Phase

Before Submitting the Proposal ▲▼

Summary Report

Answered (current phase): 50/51

Answered: 93/115

Chapters

I.	Administrative details	✓
II.	Re-using data	✓
III.	Creating and collecting data	1
IV.	Processing data	✓
V.	Interpreting data	✓
VI.	Preserving data	✓
VII.	Giving access to data	✓

More

Summary Report

Metric	Measure	
Findability	0.50	
Accessibility	1.00	
Interoperability	0.48	
Reusability	0.33	
Good DMP Practice	0.67	
Openness	1.00	

Administrative details

- ❓ Help   ›
- Rob Hooft   ›

« Collapse sidebar

DS Wizard

- Knowledge Model Editor
- Knowledge Models
- Questionnaires
- Documents
- Storage Costs Evaluator

Create document

Name

For FAIR Maturity WG

Questionnaire

For FAIR Maturity WG

Answered (current phase): 50/51

Answered: 93/115

Template

Science Europe DMP

Format

- HTML Document
- PDF Document
- LaTeX Document
- MS Word Document
- OpenDocument Text
- Markdown Document

Cancel   Create

Help

Rob Hooft

« Collapse sidebar

Section A: Data Collection

1. What data will you collect or create?

Data formats and types

We will be using the following data formats and types:

- **Powerpoint slides**

  It is a standardized format. We are aware that this is not a suitable format for long-term archiving. We will have only a small amount of data stored in this format.

2. How will the data be collected or created?

There will be no instrument dataset in this project.

Storage and file conventions

We will use a filesystem with files and folders with the following folder conventions:

- There will be a **folder for each sample/subject**. Each of those will use the

Conclusions

- DS Wizard supports FAIR-based DMPs
  - Implementation choices are guided by (predicted) FAIRness of the result
- Funders can quickly decide whether the DMP is FAIR enough
- GOAL 1: Benefit to the researcher and the project
  - Make researcher aware of technology choices and (local) services available
  - Make researcher aware of the benefits of FAIR
- FAIR evaluation
  - Not an exam, but awareness / education / checklist
  - Help make a wise decision of investment vs benefits of every choice
  - Help funders evaluate DMPs

DMPs and FAIRsFAIR

Angus Whyte, Patricia Herterich

Digital Curation Centre & FAIRsFAIR

FAIR Data Maturity WG 16 September 2020

@FAIRsFAIR_EU

Outline

- Brief recap of FAIRsFAIR project and assessment of digital (data) objects

- Interest in DMPs – motivations, recommendations

- Synergies between DMP evaluation and FAIR data maturity

- Prospective case studies and guidance

FAIRsFAIR - Fostering Fair Data Practices in Europe

https://www.fairsfair.eu

- Aims to supply practical solutions for the use of the FAIR principles throughout the research data life cycle.

- Budget: €10 million

- 22 partners from 8 member states

FAIR assessment of digital (data) objects

The indicators developed as part of the RDA FAIR data maturity model working group have the sole purpose of answering the question '**What** needs to be measured to assess the FAIRness of a digital object' and **not 'How to** measure the FAIRness of a digital object'

Research Data Alliance FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model: specification and guidelines - draft. Research Data Alliance. https://doi.org/10.15497/RDA00045
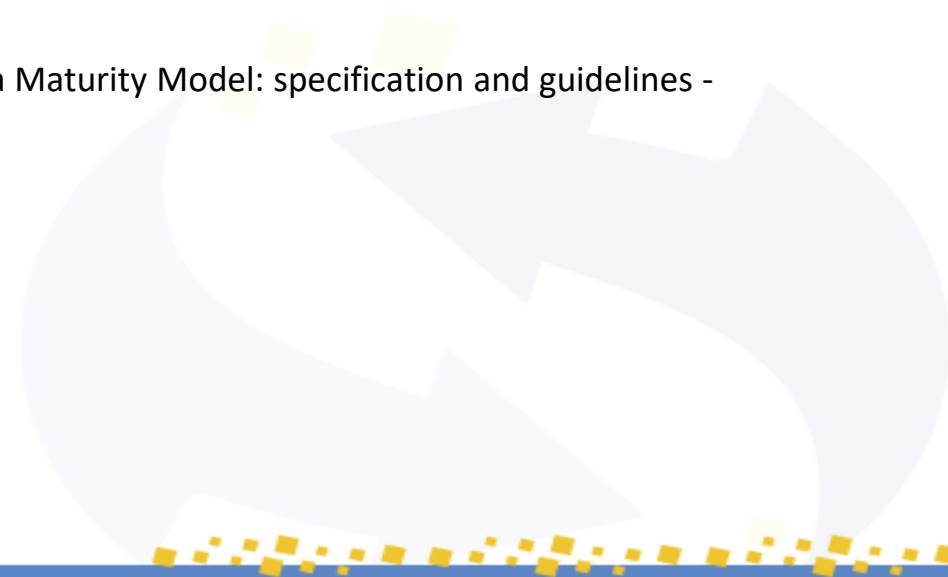
FAIRsFAIR Data Assessment Metrics

- **There are 15 metrics built on existing work.**
  - RDA FAIR Data Maturity Model
  - DANS Fairdat/FAIREnough
  - WDS/RDA Assessment of Data Fitness

- **To be piloted, next release in August**

References

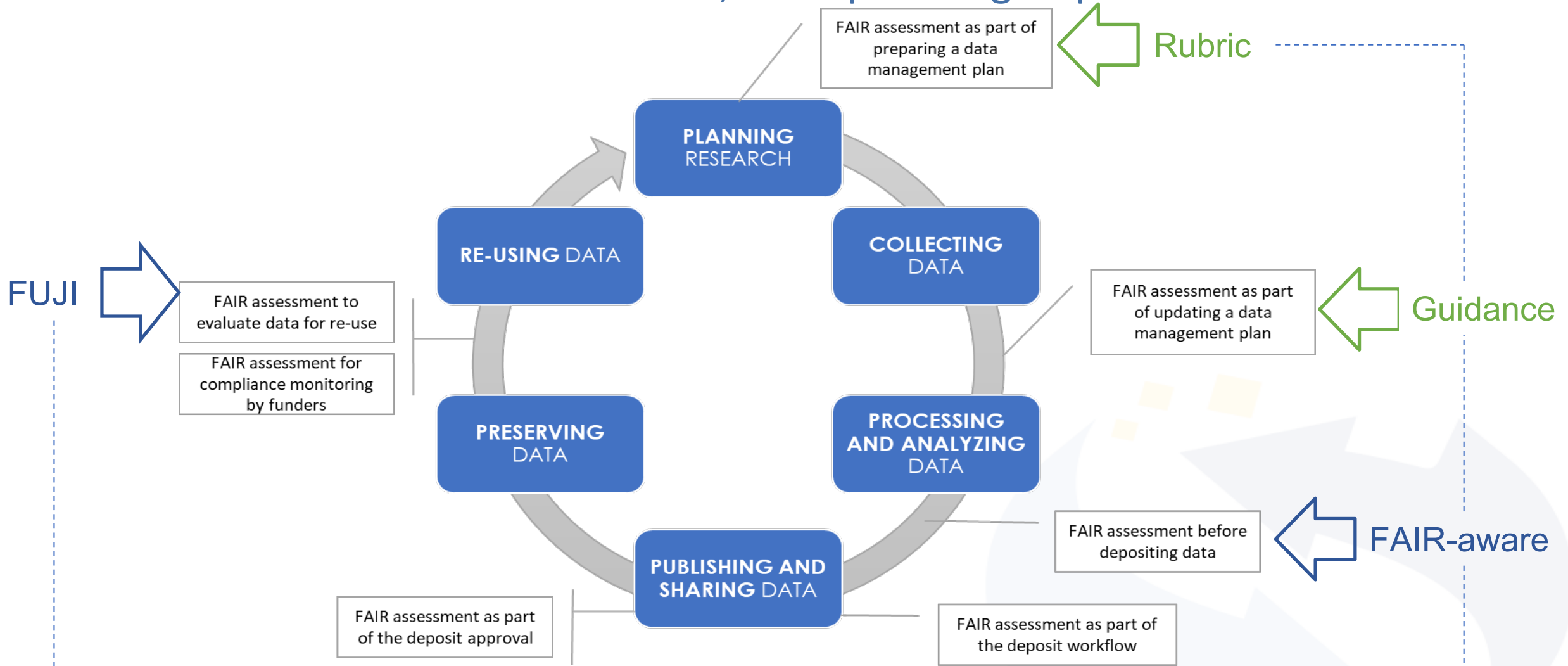D4.1: Draft recommendations on requirements for FAIR data objects in trustworthy data repositories https://doi.org/10.5281/zenodo.3678716

FAIRsFAIR Data Objects Assessment Metrics (Version 0.2) http://doi.org/10.5281/zenodo.3775794

FIELD	DESCRIPTION
Metric Identifier	FsF-F1-01D
Metric Name	**Data** is assigned with a globally unique identifier.
Description	A data object may be assigned with a globally unique identifier such that it can be referenced unambiguously by humans or machines. Globally unique means an identifier should be associated with only one resource at any time. Examples of unique identifiers of data are Uniform Resource Identifier (URI) such as URL and URN, Digital Object Identifier (DOI), the Handle System, identifiers.org, w3id.org and Archival Resource Key (ARK). A data repository may assign a globally unique identifier to your data or metadata when you publish and make it available through their services.
FAIR Principle	F1. (Meta) data are assigned globally unique and persistent identifiers
CoreTrustSeal Alignment	R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
ASSESSMENT	
Requirement(s)	• Data identifier (IRI, URL) • List of globally unique identifier schemes
Method	Check if the data identifier is specified based on a globally unique identifier scheme.
COMMENTS	
Related Resources:	

FAIR assessment should be iterative, from planning to publication

FAIR-Aware(left) https://fairaware.dans.knaw.nl/ and automated assessment tool FUJI (right) https://github.com/pangaea-data-publisher/fuji

DMP assessment should be iterative, from planning to publication

Recommendations to support FAIR Data Principles

> "(14) Working with all stakeholders, ensure that data management planning is supported across the entire research lifecycle so that data can be "born FAIR" and kept "FAIR enough" over time. Require updating of DMPs over the research lifecycle leading to comprehensive, high quality end-stage DMPs that are included in end-stage reporting.
>
> (16) Support researchers to assess the potential risks, benefits and associated costs to enable the sharing of FAIR data as they draft their DMP. "

D3.3 Policy Enhancement Recommendations
https://zenodo.org/record/3686901

Iteratively assess plan on whether it includes appropriate actions to make data FAIR

Research communities, funding bodies, publishers, RIs, RPOs, associations, membership organisations, and partnerships

Recommendations to support FAIR Data Principles

> B1: Formalise and support appropriate data management plans (DMPs) for FAIR data
> - Researchers, data stewards and academic journals can use FAIR data criteria to review whether a DMP includes appropriate steps to make the data FAIR.
>
> B2: Develop roadmaps, guidance and workflows for machine-actionable DMPs to inform FAIR data stewardship
> - DMP platform providers, institutions, repositories and other data service providers should adopt the RDA Common Standard for Machine-Actionable Data Management Plans to exchange DMP content
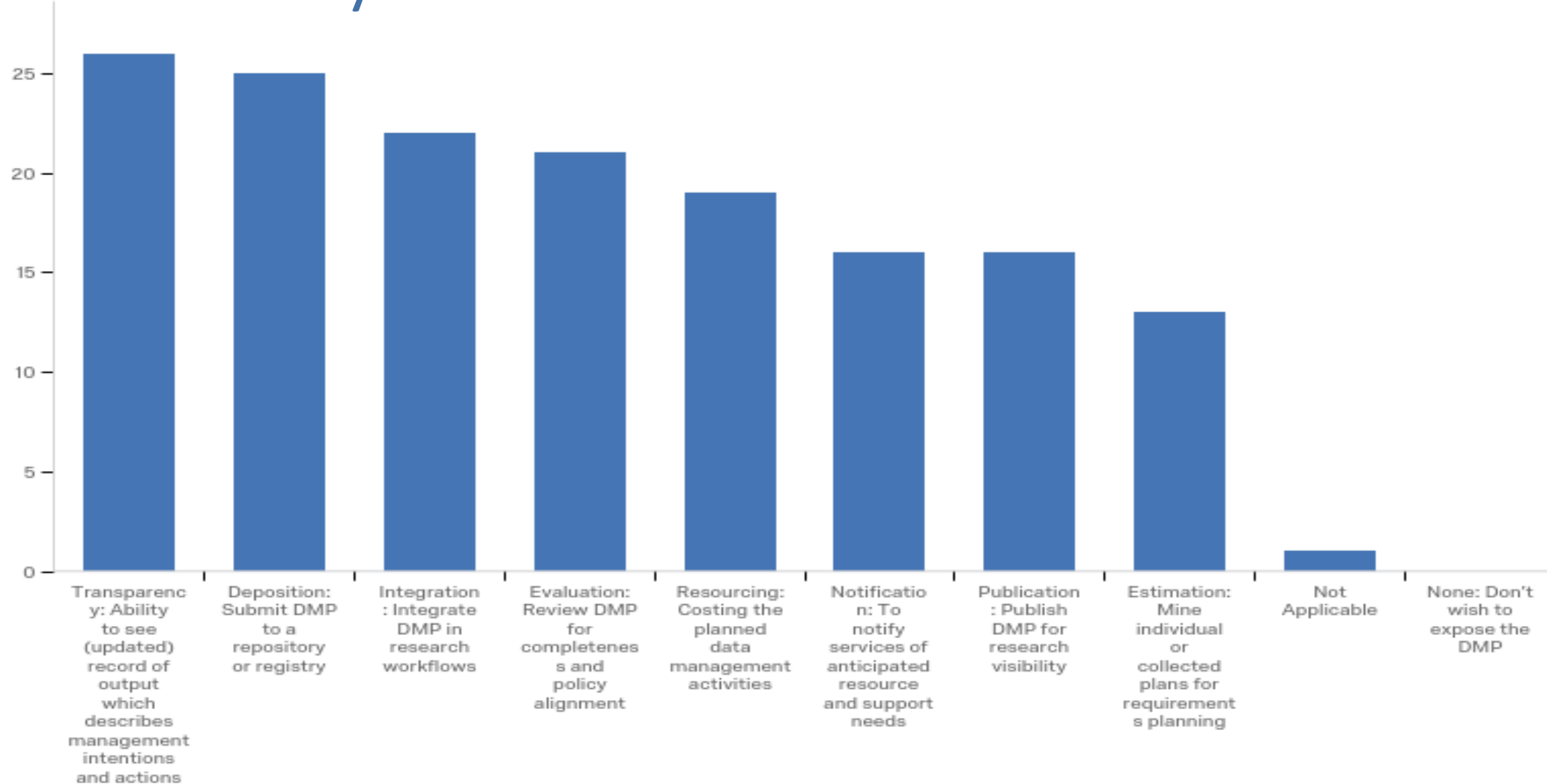
Make machine-actionable DMPs serve FAIR data assessment

Research communities, data stewards, repositories, other data service providers and journals

D3.4 Recommendations on Practice to Support FAIR Data Principles
https://zenodo.org/record/3924132

Supporting actions…

1. Publish case studies on selected maDMP **use cases** in collaboration with others e.g. funders, GO-FAIR, RDA Working Groups

2. Co-create demonstrators using FAIR indicators in DMP guidance and assessment to evaluate mutual benefits to repositories and research communities

Community interest in maDMP use cases

N= 571 responses to RDA Exposing DMP Survey

Demonstration

Investigating use cases

- Present FAIR guidance in DMP tools

- Assess how FAIR dataset will be from a machine actionable DMP, adapting the metrics applied by FUJI

```json
[
  {
    "id": 71,
    "title": "Sample plan",
    "grant_number": null,
    "last_updated": "2019-10-15 22:30:42 UTC",
    "creation_date": "2019-10-02 22:21:30 UTC",
    "template": {
      "title": "NSF-GEN: Generic",
      "id": 187
    },
    "funder": {
      "name": "National Science Foundation (NSF)"
    },
    "principal_investigator": {
      "name": null,
      "email": null,
      "phone": null
    },
    "data_contact": {
      "name": null,
      "email": null,
      "phone": null
    },
    "users": [
      {
        "email": "tester@example.org"
      }
    ],
    "description": null,
    "plan_content": [
      {
        "title": "Data Management Plan",
        "description": null,
        "sections": [
          {
            "title": "Types of data produced",
            "description": null,
            "number": 1,
```

DMPonline

- FAIRsFAIR guidance

Thank you for your attention!

Slide acknowledgements:

Ilona von Stein (DANS)

Anusuriya Devaraju (PANGAEA)

Herve L'Hours (UKDA)

Mustapha Mokrane (DANS)

www.fairsfair.eu

@FAIRsFAIR_EU

OPEN

DISCUSSION

FAIR use cases

Funding agencies representatives are going to present the extent to which they are involved in **FAIR assessment** and in what context.

10' — FAIRware – a software tool to assess FAIRness

Adam Dinsmore
Wellcome

10' — FAIR requirements under Horizon Europe
Carlos Casorrán & Konstantinos Repanas
DG RTD

10' — Open discussion
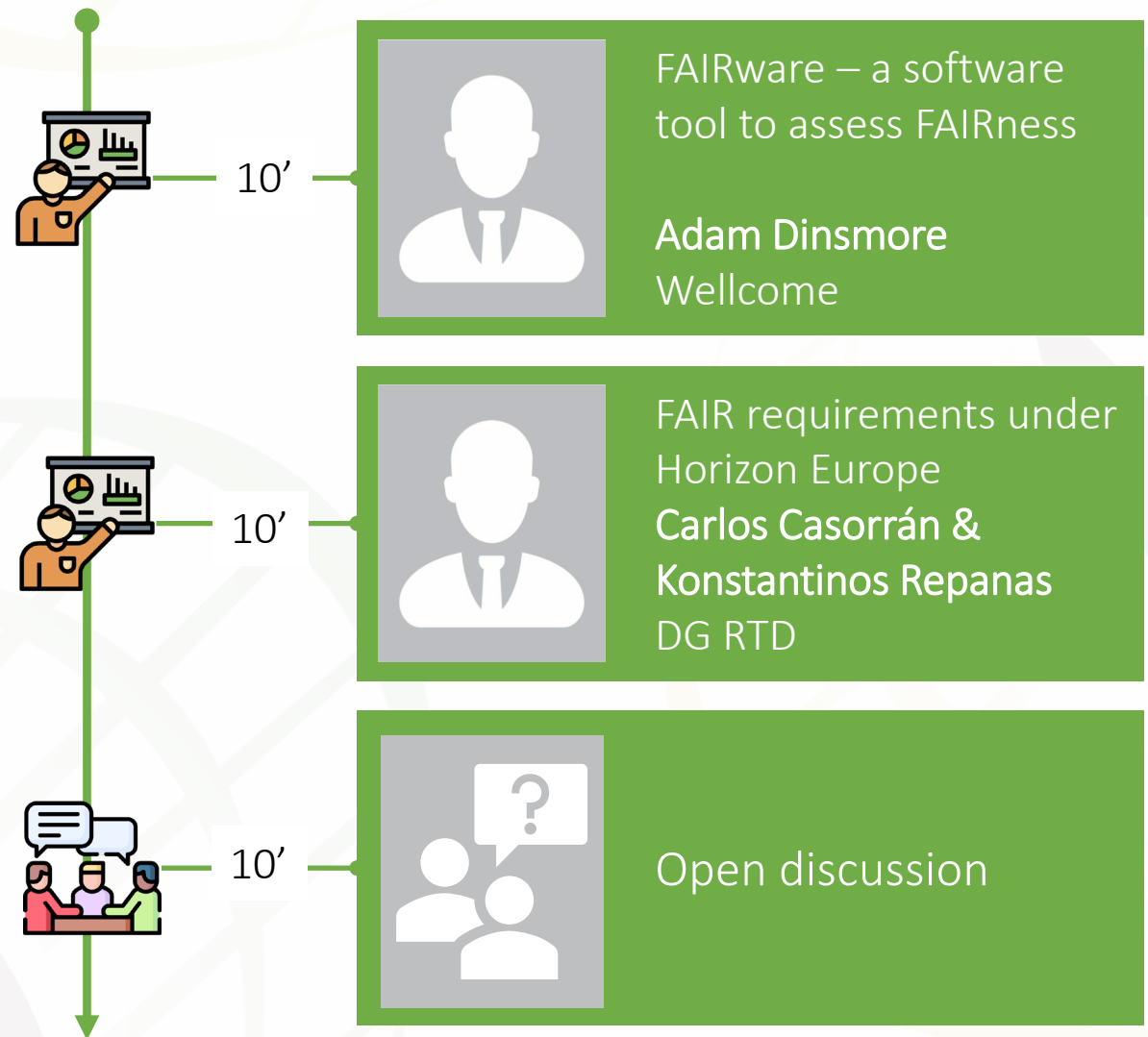
FAIRware:
Designing and embedding open source software to support researchers and funders in implementing the FAIR principles

DAVID CARR, WELLCOME TRUST; ADAM DINSMORE, RORI

RoƦI RESEARCH ON RESEARCH INSTITUTE

The rationale for FAIRware

FAIRware: What is the need?

- RoRI is committed to open science and maximising the value of research outputs – including publications, datasets, & software.
- Many RoRI Partners have long-standing policy requirements in place requiring funded researchers to:
  - ensure open access to all research articles, monographs and book chapters.
  - maximise the availability of research data and software with as few restrictions as possible.
- Many funders are increasingly encouraging their researchers to make research data and other outputs available in line with the FAIR principles.
- FAIRware brings funders together to drive consensus and synchronise efforts in this space.

RoRI RESEARCH ON RESEARCH INSTITUTE

Automating assessment of FAIR
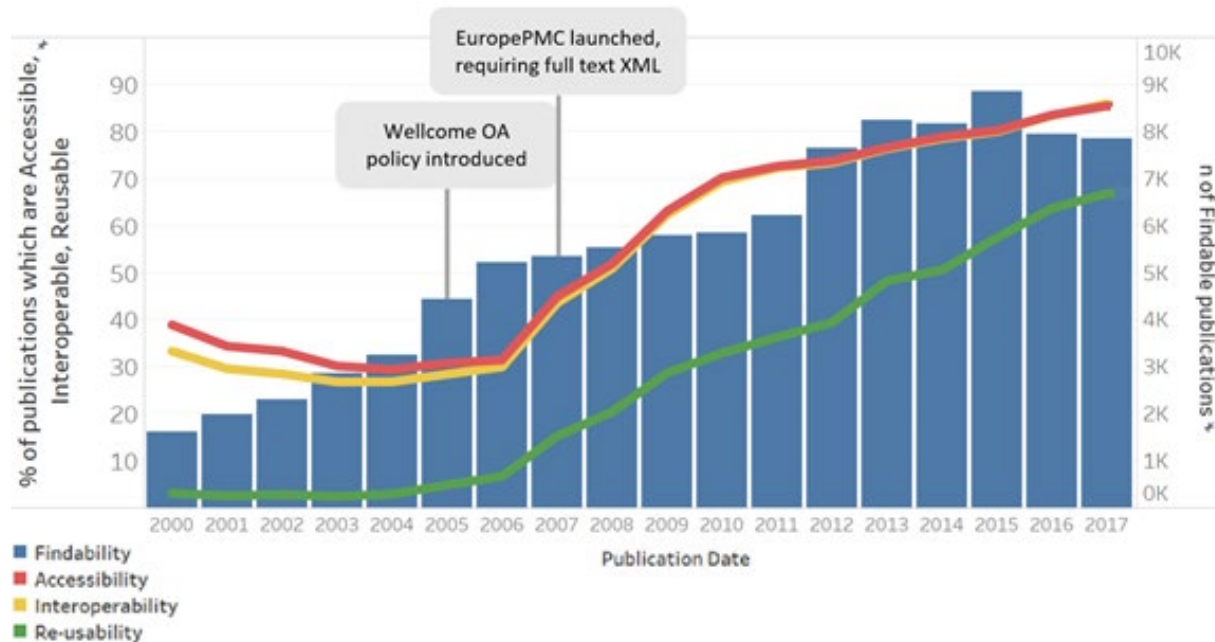
- Currently possible to monitor and evaluate FAIRness of research articles with high fidelity.

- Possible because metadata provenance high for articles.

- However, at present we are unable to access similar data – and therefore think strategically – about other output types.

- Guidance for researchers is dispersed, variable, manual.

- In part because of fuzziness around what the FAIR principles mean in practice, lack of consensus among funders.

Keyword(s)	Title	Timetable	Strategic Partners involved / interested	Core team leads	Next steps
CRITERIA	**How do grant application criteria influence inequalities in research funding?**	June 2020 - Sept 2021 (+ possible qualitative follow-up project)	**11 partners as of Sept 2020:** AAS. CIHR; CZI; EMBO; MSFHR; Novo Nordisk Fonden; NRC; NWO; W/DBT India Alliance; UKRI; Wellcome	**Matt Thakur** Vincent Traag Danil Mikhailov Wolfgang Kaltenbrunner	Initial scoping work now underway. Project partner meetings in late May/ early June
EXCELLENCE	**Transforming 'excellence': definitions, applications and alternatives to excellence in the evaluation, management & measurement of research**	July 2020- Sept 2021	**11 partners as of Sept 2020:** AAS, ARC; CIHR; CZI; EMBO; FWF; HHMI; MSFHR; NIHR; SNSF; Wellcome	**Stephen Pinfield** **Thomas Franssen** Sarah de Rijcke	Initial scoping now underway. Project meetings in late May/early June. Drop-in session 12:00-13:00 BST Weds 27 May
FAIRware	**FAIRware: an open source software tool to assess the FAIR status of research outputs**	June 2020 – Sept 2021	**5 partners as of Sept 2020:** CIHR; FWF; NIHR; SNSF; Wellcome	**Adam Dinsmore** Danil Mikhailov	Work underway. Request for Proposals published in September 2020.
PATHWAYS	**Career pathways in research: a six-country, multi-partner study of data, frameworks & future directions**	July 2020- Sept 2021 (+ possible further phase into 2022)	**11 partners as of Sept 2020:** AAS, CIHR; CZI; FWF; HHMI; MSFHR; Novo Nordisk Fonden; Sloan; UKRI; Volkswagen Foundation; Wellcome	**Anne-Marie Coriat** James Wilsdon Inge van der Weijden	Scoping underway. Partner cluster or 1-1 calls to be arranged for early June.
RANDOMISATION	**Focal randomisation in grant funding: a multi-funder comparative study**	July 2020- Sept 2021 (+ possible expansion into further funder experiments)	**15 partners as of Sept 2020:** ARC; CIHR; CZI; EMBO; FWF; MSFHR; NIHR; Novo Nordisk Fonden; NWO; NRC; Sloan; SNSF; UKRI; Volkswagen Foundation; Wellcome	**James Wilsdon** **Sarah de Rijcke** Ludo Waltmann Helen Buckley Woods	Scoping underway. Partner cluster or 1-1 calls to be arranged for early June.

RoRI projects – overview of first wave (2020-2021)

RoRI RESEARCH ON RESEARCH INSTITUTE

FAIRware scoping to date

- **Request for Information** sought feedback from potential suppliers and other organisations active in the FAIR data space.
  - Responses indicated that the proposed initiative was **feasible and timely**.
  - **Cultural challenges** remain i.e. consensus building among Funders w/r/t the FAIR principles.
  - **Many groups already active** on the problem of defining FAIR and making it machine actionable.
    - FAIRsFAIR.eu
    - Several RDA Working Groups (inc. FAIR Data Maturity Model).
    - FAIRSharing.org.
    - GO FAIR (inc. M4M workshops).

- **RoRI well placed to act**, drawing together:
  - Material resource to support software development.
  - Technical capacity to lead project.
  - Convening power to draw funders together to build the necessary consensus.

Project Governance Structure

Steering Group

5 partners as of June 2020:
CIHR; FWF; NIHR; SNSF; Wellcome

Advisory Partner

Project Group

Wellcome Data Labs

+

Development Partner(s)

FAIRware developer community.
Members to be selected by RFP (December 2020).

Project partner – FAIRsFAIR.eu
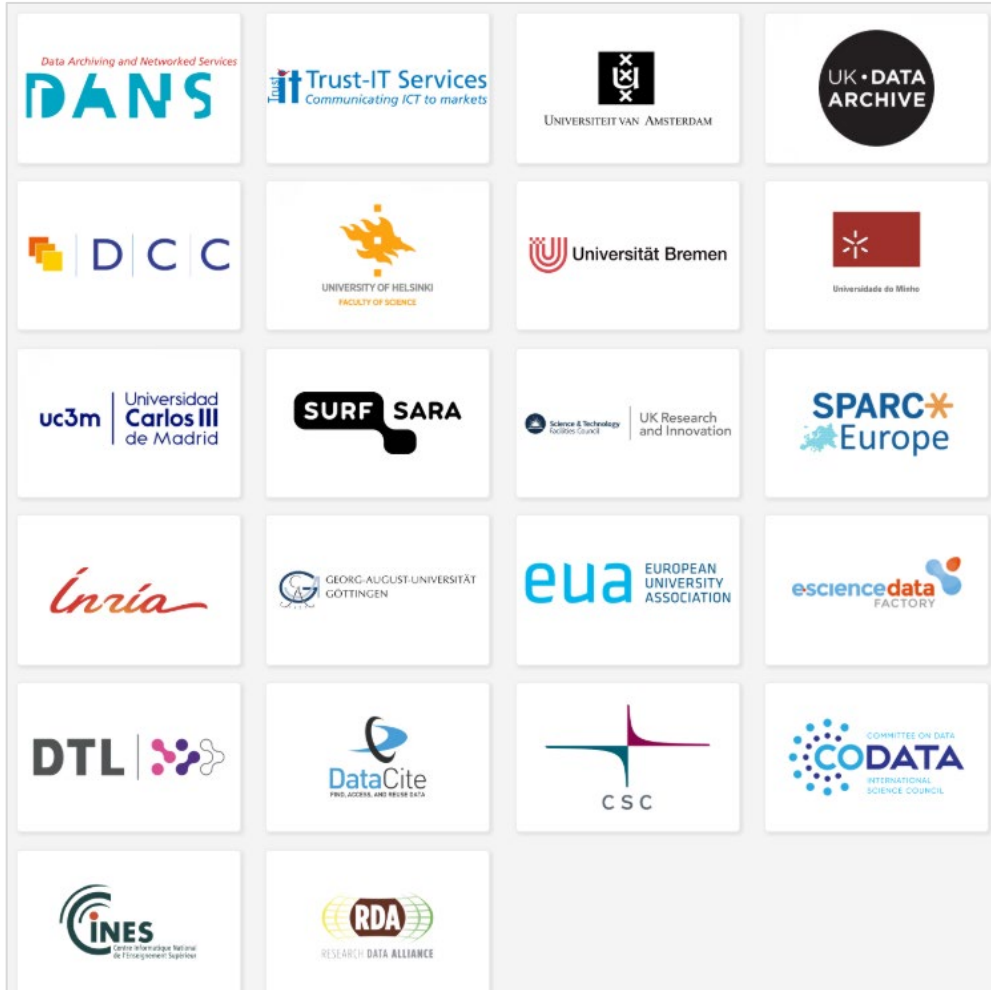
- **FAIRsFAIR** is a multi-centre collaborative programme funded by the European Commission and coordinated by DANS.

- FAIRsFAIR aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle, and embed FAIR principles and practices as a key element of the European Open Science Cloud.

- Will advise the FAIRware Steering/Project groups on implementation of FAIR assessment for research data and software (subject to MOU).

RoRI RESEARCH ON RESEARCH INSTITUTE

FAIRsFAIR Data Object Assessment Metrics

Figure 1. Anatomy of FAIRsFAIR metric identifier.

The following is a list of 13 FAIRsFAIR data assessment metrics. At present, the metrics address the FAIR principles, except A1.1, A1.2 (communication protocol) and I2 (FAIR vocabularies).

Table 2. List of Metrics.

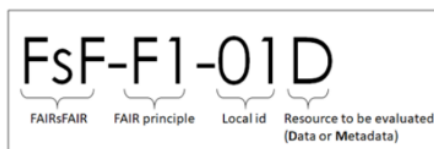| Identifier	Name
FsF-F1-01D	**Data** is assigned a globally unique identifier.
FsF-F1-02D	**Data** is assigned a persistent identifier.
FsF-F2-01M	**Metadata** includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.
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.

FIELD	DESCRIPTION
Metric Identifier	FsF-F1-01D
Metric Name	**Data** is assigned a globally unique identifier.
Description	A data object may be assigned with a globally unique identifier such that it can be referenced unambiguously by humans or machines. Globally unique means an identifier should be associated with only one resource at any time. Examples of unique identifiers of data are Internationalized Resource Identifier (IRI)[12], Uniform Resource Identifier (URI) such as URL and URN, Digital Object Identifier (DOI), the Handle System, identifiers.org, w3id.org and Archival Resource Key (ARK). A data repository may assign a globally unique identifier to your data or metadata when you publish and make it available through its curation service.
FAIR Principle	F1. (Meta) data are assigned globally unique and persistent identifiers
CoreTrustSeal Alignment	R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation
ASSESSMENT	
Requirement(s)	• Data identifier (IRI, URL) • List of globally unique identifier schemes
Method	Check if the identifier is specified based on a globally unique identifier scheme.
COMMENTS	

Related Resources:
• Identifiers compiled by FAIRsharing, https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema
• A list of Uniform Resource Identifier (URI) schemes, available in different formats, https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml#uri-schemes-1
• Uniform Resource Identifier (URI) Generic Syntax (RFC 3986), https://tools.ietf.org/html/rfc3986

FAIRsFAIR Data Object Assessment Metrics:
https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments

RoRI RESEARCH ON RESEARCH INSTITUTE

FAIRware RFP (e.t.a. Sept 2020)

- **Two Key Tasks**: applicants will be asked to describe how they would 'DESIGN' and/or 'BUILD' a MVP in partnership with Wellcome Data Labs as part of an Open Source Community of Collaboration

- Applicants can bid to undertake one or both key tasks.

- Joint applications from teams with complementary skills welcomed.

- FAIRsFAIR an official project partner.

- Applicants encouraged to challenge our thinking, suggest alternative approach.

https://wellcome.ac.uk/what-we-do/our-work/research-research

RoRI RESEARCH ON RESEARCH INSTITUTE
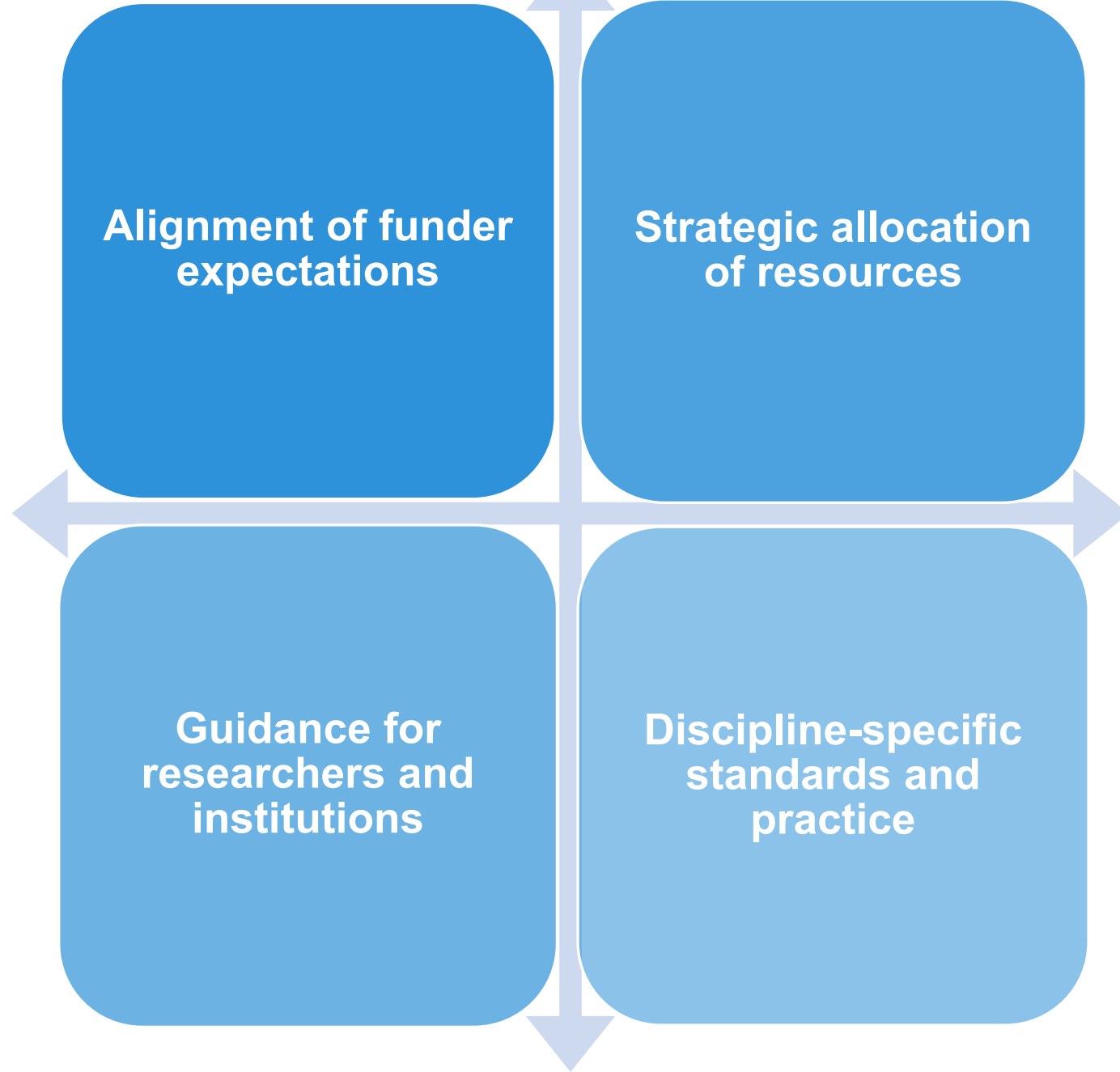
Key Design Principles

- **Open source:** all code must be open source, so it can be assessed, built upon and used by the community to enhance the tool.

- **Accessibility:** the tool must meet WCAG 2.1's AA standards for web content accessibility.

- **Harmonisation:** the tool should build upon existing and ongoing initiatives in the FAIR data space.

- **Modular, iterative, adaptable:** as FAIR policies and guidelines develop, FAIRware will need to develop in lockstep to ensure the assessments it provides remain relevant to user requirements.

- **Breadth of application:** the tool should be based on core criteria for assessing the FAIRness of particular output types that are applicable across the full spectrum of research fields but allow for the incorporation of discipline-specific standards where they exist and over time.

- **Partnership and buy-in:** the consensus criteria on which the tool is based should have the buy-in and support of partner funders and the wider FAIR data community.

- **User focus:** the development of the tool should incorporate extensive usability testing to ensure that it meets the needs of key user groups (funders, researchers and institutions).

- **Actionable advice:** the tool should provide practical and actionable advice on how the FAIRness of outputs may be enhanced.

- **Sustainability:** the tool must be actively maintained and developed over time so its value persists, with applications encouraged from devs future-proof tech stacks.
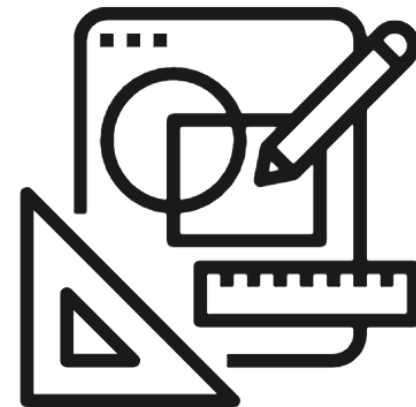
Driving the paradigm shift towards Open Science: FAIR data

Kostas Repanas & Carlos Casorrán

European Commission

Directorate-General for Research and Innovation

Open Science unit

RDA's FAIR Data Maturity Model Workshop

Sept 16th, 2020

Data

- Data is the **main asset** of the digital economy

- Data production continues to **grow exponentially**→ from 33 zettabytes in 2018 to 175 zettabytes in 2025

European Data Market

Data workers
6.16 million in 2016
↳ 10.43 million by 2020

Data companies
255,000 in 2016
↳ 359,050 by 2020

Data economy value
Almost € 300 billion in 2016 → € 739 billion by 2020

European Commission | Source: European Data Market study
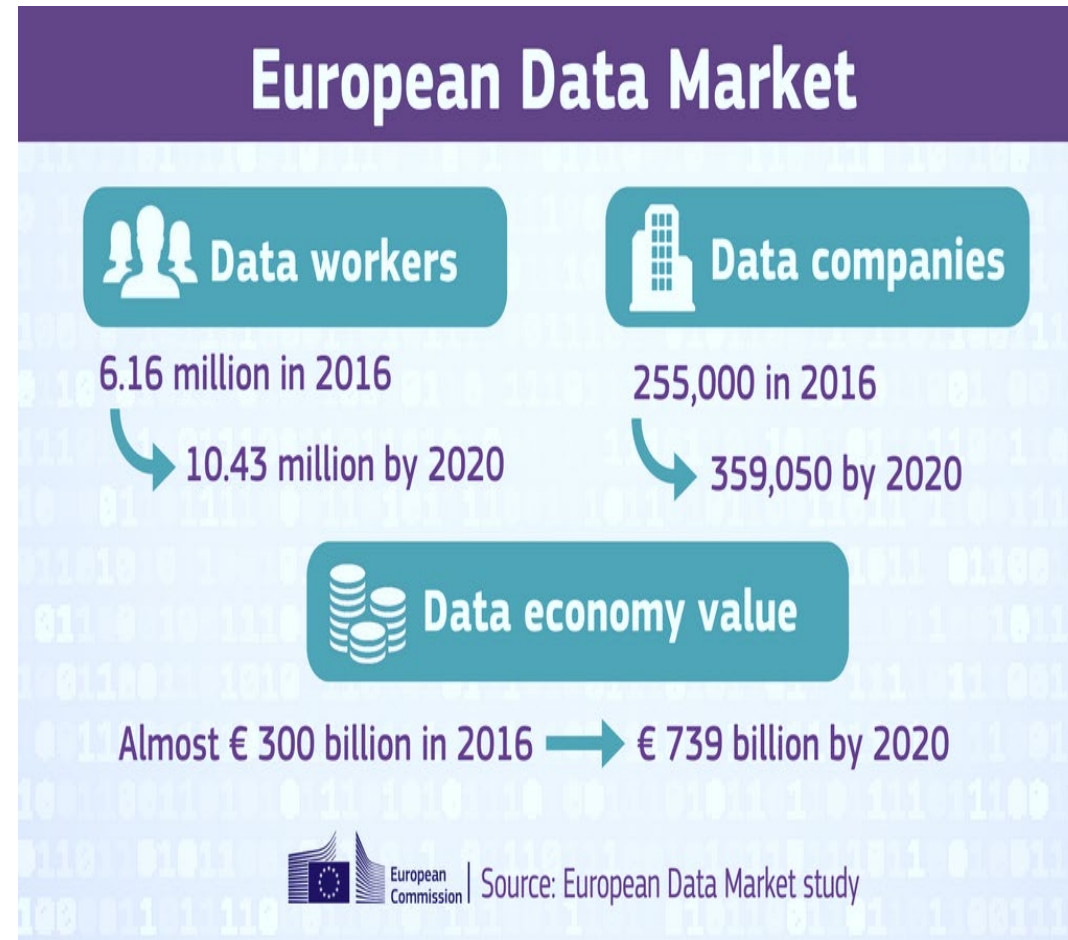
- Poor data management incurs an **opportunity cost** in the billions of euros

- **Research/Science** is one of the main data producing/consuming domains

49

Open Science

- **Open Science** means **sharing knowledge and tools as early as possible**, not only **between researchers** and **between disciplines**, but also with **society at large**.

- Open Science improves the **quality**, **efficiency** and **creativity** of research and the **trust by society in science**. In particular, OS is beneficial for science, scientists and funders, e.g.:

  - tackles the reproducibility crisis;

  - faster response to societal challenges e.g. Coronavirus, Ebola;

  - access to and sharing results yields higher impact through collaborations;

  - generates new research findings;

  - large opportunity costs of non-FAIR data—€10.2bn/year (source: Cost-benefit analysis of FAIR research data, 2017).

- The Commission acts as **policy maker** (propose legislation and encourage MS), a **funder** (we set requirements to our projects) and a **capacity builder** (we fund 'enabling' projects).

European Commission

Horizon 2020 & Horizon Europe

Horizon 2020: ~ €80bn

Industrial Leadership
EUR 17.0 billion

Societal Challenges
EUR 29.7 billion

Excellent Science
EUR 24.4 billion

European Institute of
Innovation and Technology
EUR 2.7 billion

Euratom
EUR 1.6 billion

Other
EUR 3.2 billion
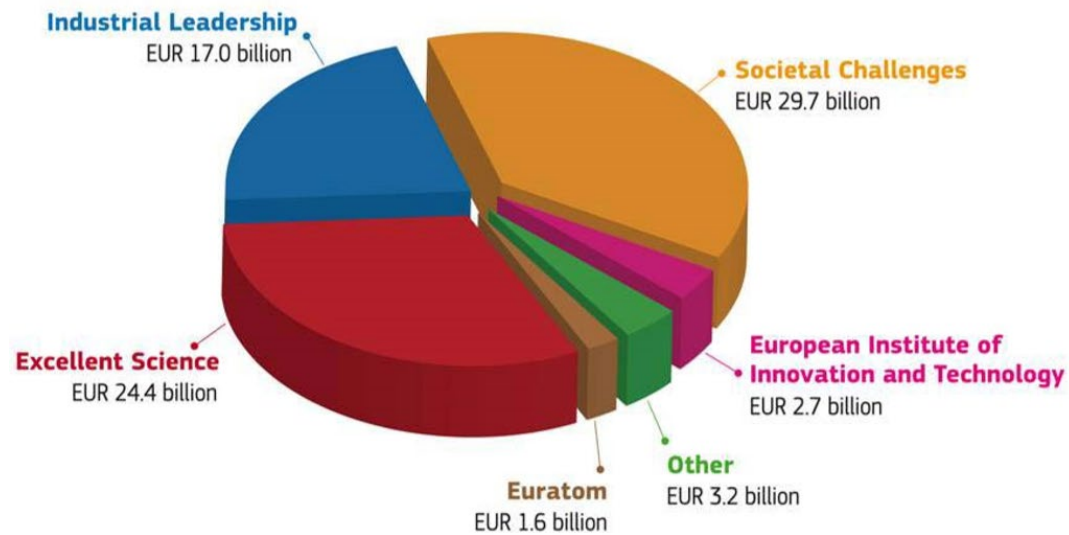
Horizon Europe: ~€100bn

€2.1 €2.4

€13.5

€25.8

€52.7
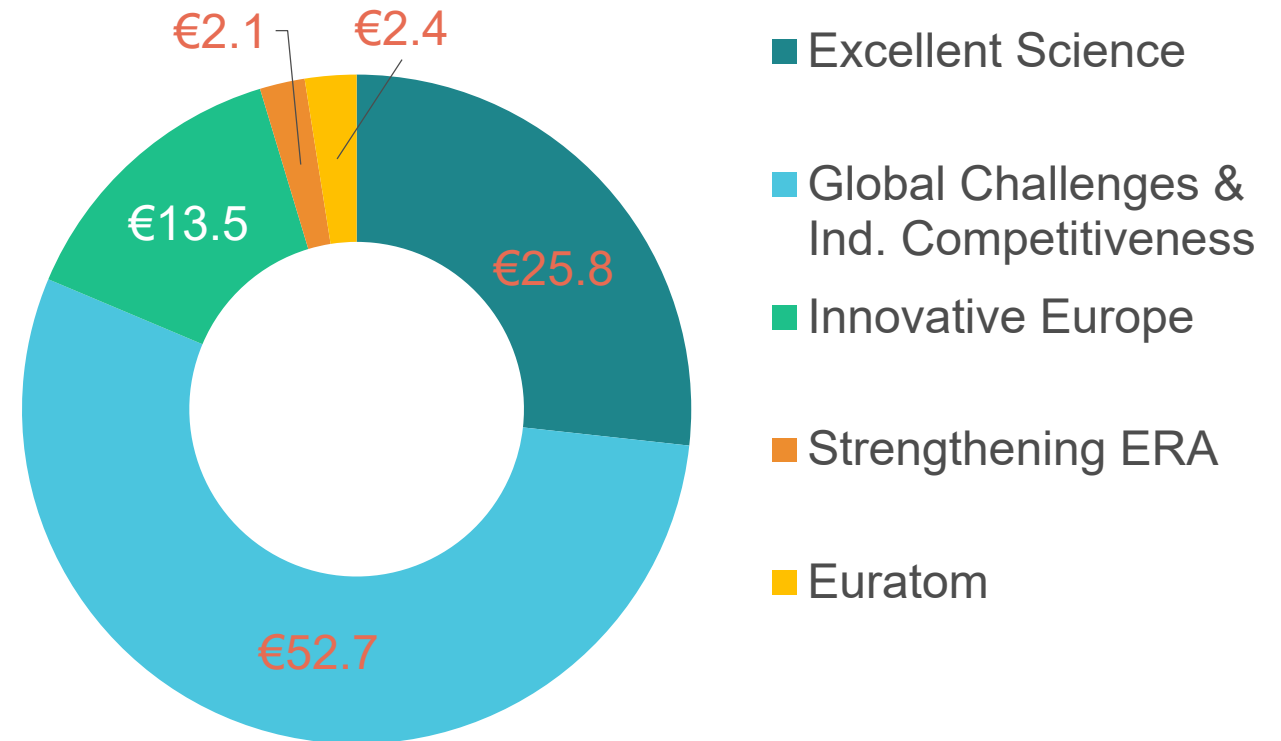
- Excellent Science
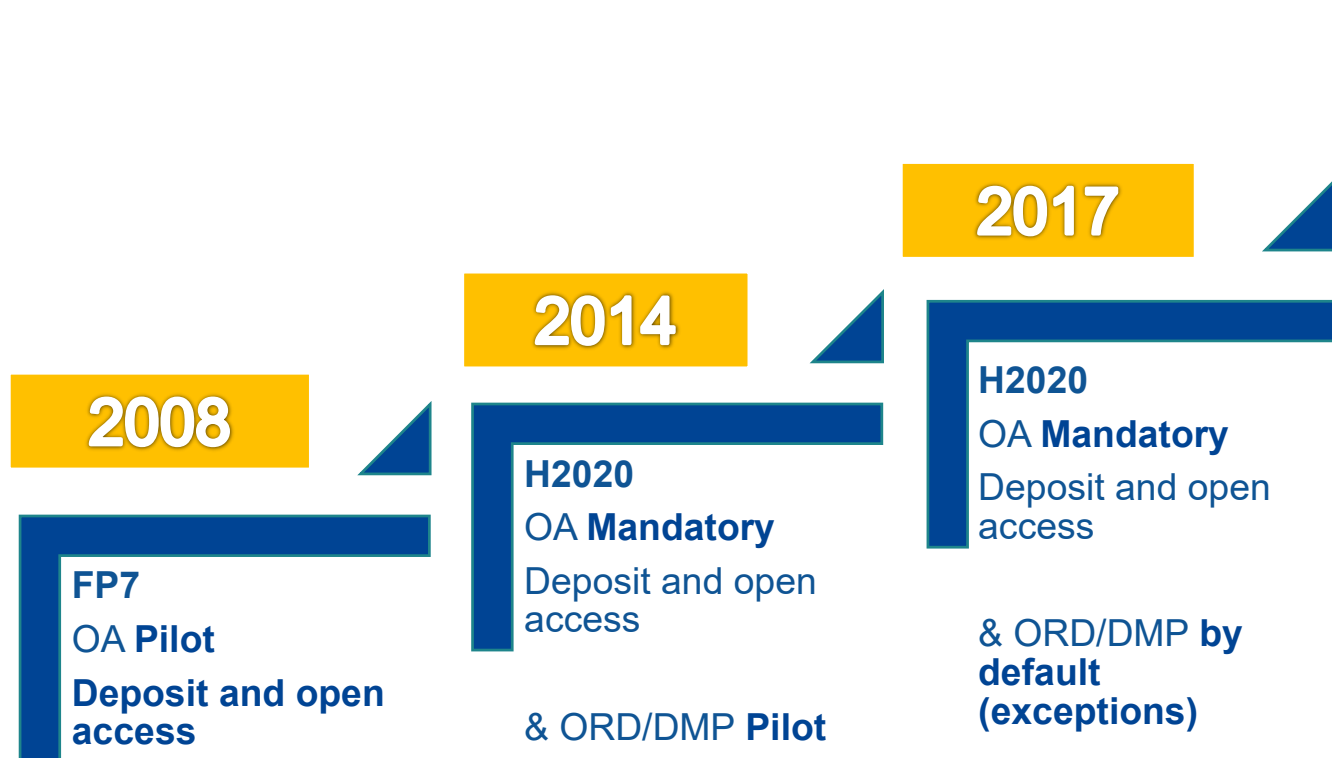- Global Challenges & Ind. Competitiveness
- Innovative Europe
- Strengthening ERA
- Euratom

- The Commission invests heavily in Research and Innovation.
- Over **30000 H2020 projects**—Projects produce **research outputs**, **data**, **deliverables**, etc.
- It becomes increasingly important to make the **best possible use** of previous work.

European Commission

Evolution of our policies across the FPs

2008

FP7
OA **Pilot**
Deposit and open access

2014

H2020
OA **Mandatory**
Deposit and open access

& ORD/DMP **Pilot**

2017

H2020
OA **Mandatory**
Deposit and open access

& ORD/DMP **by default (exceptions)**

We are pioneers!

Planned under **Horizon Europe (2021)**:

- Open Science (OA, RDM, Cit. Eng., etc.) embedded throughout the FP. OS to play a role in the:
  - **Evaluation** of proposals (methodology)
  - **Grant Agreement**
  - **Reporting**—during the project's lifetime

- Strengthening of the obligations with respect to open access and focus on responsible RDM in line with FAIR

European Commission

Research data under Horizon Europe

- The governing principle will be to manage research data **responsibly**, **in line with FAIR**, and for data to be "**as open as possible and as closed as necessary**". To accomplish this, we are **likely** to see the following:

  - At proposal stage, beneficiaries will be **evaluated** on **preliminary RDM considerations**

  - All projects that **generate/collect/re-use** research data will have to establish and regularly update a **Data Management Plan** (living document and machine-actionable)

  - Beneficiaries will have to **deposit data in a repository** and link their data to publications they underpin, if applicable

    - Valid repositories will be those that provide **PIDs** for the data and ensure **rich metadata** in line with FAIR.

    - For some actions, an additional obligation to deposit in a repository that is federated under EOSC.

  - Beneficiaries to ensure **open access ASAP** under CC BY or CC0 licences (or equivalent), unless exceptions apply (duly justified in the DMP)

Other research outputs under Horizon Europe

- Data is not the only research output that should be managed → **digital**: Software, algorithms, protocols, workflows, models, and **physical**: reagents, antibodies, hardware, etc. all need to be properly managed:

  - We would like to see other research outputs also **described in the DMP**.

  - There will be a **strong encouragement** for beneficiaries to **deposit and provide open access via a repository** to these outputs, unless legitimate interests or constraints apply.

- Research outputs, either digital or physical, should also be findable, interoperable, accessible and re-usable → We need a better understanding of

  - What **FAIR** means for digital objects, other than data (e.g. software),

- A better understanding of **FAIRness beyond data**, will likely **inform future funder requirements** aimed at, **addressing the reproducibility crisis** and **maximizing the impact of new research** by promoting and facilitating building on previous research.

European Commission

FAIR and data sharing against the Sars-CoV-2

- This effort represents a practical use case of **FAIR data in action**.

- The European Commission launched on 20 April the **European COVID-19 Data Platform** together with EMBL-EBI, ELIXIR, and other partners, and with the support of MS.

- The Platform is a thematic priority pilot to realise the **EOSC vision** and to showcase the added value of **FAIR data sharing** to advance science and benefit researchers

- It responds to the need to capitalise on the quick and wide **sharing**, **re-use**, **processing** of and **access** to **research data** and **metadata** on the SARS-CoV-2, and the related COVID-19 disease.

- A very strong focus is placed on ensuring that data and metadata on this Platform are as **open** and as **FAIR** as possible (not always possible, because of the types of data, or because of the limited curation at the submission end). To maximise the FAIRness of the data, e.g. ELIXIR-NO is performing manual curation on the viral sequences received.

- Crucial lesson: Researchers need **quick** and **unrestricted** access to multiple data sources to accelerate their research. **FAIR data** is an essential component in this.

How do we move forward?

- **FAIR data** are an essential component in **Open Science policies** to use the wealth in data to generate new and impactful research and tackle the reproducibility crisis

- As a funder, the European Commission, will impose **stringent requirements** on all Horizon Europe projects with respect to publications, data, and other research outputs

- Specifically, with respect to data, the focus will be on **responsible research data management in line with the FAIR principles**

- The community-backed work developed under the **FDMM WG** represents a blueprint on which further **evaluation mechanisms of FAIRness** can be built.

- **Assessment of FAIRness**, either by funders, or supported by FAIR-enabling repositories (certified or trustworthy), will raise awareness and increase levels of compliance across the board.

European Commission

Keep in touch

ec.europa.eu/

europa.eu/

@EU_Commission

@EuropeanCommission

European Commission

europeancommission

@EuropeanCommission

EUTube

EU Spotify

European Commission

Thank you

European Commission

OPEN
DISCUSSION

Survey on FAIR assessments

Problem statement

Improve the understanding of benefits and challenges of the FAIR assessments from the perspective of the funders and communities.

Objectives

1. Formulate conclusions and recommendations on the level of policy, (i.e. better understanding of the perspectives of both sides)

2. Finding out how the research community and the funders' community might want to use the model and what changes they would want to see

Format and output

- E-survey
- Voluntary basis
- 8-point questionnaire

Anonymise, compile and analyse the results, draw conclusions and propose recommendations for further actions to improve understanding

Possible topics for the next versions of the model

- Approaches toward evaluation of FAIR assessment tools and services, taking into account community aspects

- Consensus within the communities for the priorities of the indicators, respecting the different targets for FAIRness and speeds of implementation

- Cross-community interoperability and evolution of standards

- Role of landing pages and human readable documentation

- PID practices across communities (identifiers for metadata, data, separately, combined)

- Metadata practices:

  - Role of generic platforms / repositories in improving domain-specific metadata

  - Metadata at several levels (collection level, individual level)

  - Consensus on minimum level for 'Rich' metadata

  - Shared understanding of knowledge representation

- Data granularity (collection, dataset, data item)

Quick poll – *topic prioritisation*

What should be our **top priority** for the next revision of the model?

multiple answers allowed

Join at
slido.com
#52695

Maintenance and governance

2020

Preparation for the maintenance

- Identify topics that need clarification and consensus
- Address the governance and maintenance practical aspects

2021

Maintenance and preparation for the 1st revision

Develop the topics identified and propose consensus-driven solutions to optimize the model and move away from a fit-for-all to a tailored solution

2022

1st revision of the model

Put into motion the necessary changes identified

Next steps

End of November

CODATA FAIR symposium
Wrap-up of the FAIR data
maturity model and defining
priorities for 2021

16 of September

WORKSHOP #10

survey on benefits and challenges of FAIR assessments

maintenance and governance plan

Late October – Early November

WORKSHOP #11
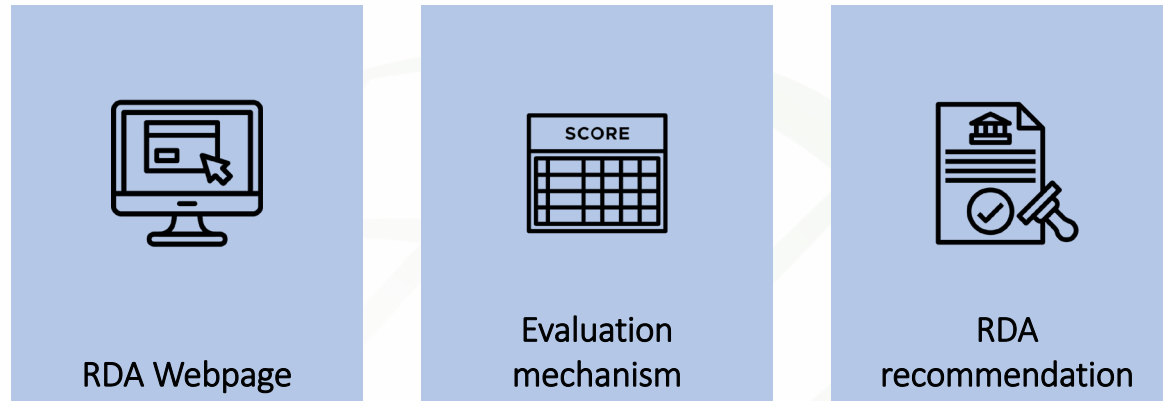Report on the survey,
developement of the topics
proposed for voting

FAIR data maturity model

Stay in touch!

https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

RDA Webpage	Evaluation mechanism	RDA recommendation

RDA FAIR data maturity model WG – GitHub

RDA FAIR data maturity model WG – Mailing list

images: Flaticon.com