RDA Global Adoption week

15 - 19 June 2020

- ➔ The RDA Global Adoption Week: 15-19 June 2020
- ➔ focused on **five areas of the research data lifecycle**

Day & Topic	Sessions
Monday, 15th June 2020 - Data Management Planning	14:00 UTC + 23:00 UTC
Tuesday, 16th June 2020 - Data Description	06:00 UTC + 14:00 UTC
Wednesday, 17th June 2020 - Identify, Store and Preserve	07:00 UTC + 14:00 UTC
Thursday, 18th June 2020 - Disseminate, Link and Find	07:00 UTC + 12:00 UTC
Friday, 19th June 2020 - Policy, Legal Compliance and Capacity	05:00 UTC + 13:00 UTC

Originally planned for the RDA 15th Plenary, the Adoption Week aims to **demonstrate the wide variety of RDA adoptable and adopted solutions to data sharing challenges** across research practices, domains and geographies.

Purpose of the week:

- Learn about RDA Outputs
- Converse with speakers from all around the world who have created and implemented them
- Determine how best to integrate those data sharing solutions into your own projects

Recommendations and outputs catalogue

- RDA Outputs are classified as **RDA Recommendations** *(official, endorsed results of RDA Groups),* **Supporting Outputs** *(useful solutions from our RDA Working and Interest Groups)* or **other Outputs**
- **They can be searched according to their status, Data Lifecycle topics or scientific domain**

rd-alliance.org/recommendations-and-outputs/catalogue

Tell your adoption story

- **Are you an adopter?** RDA is actively seeking new adoption stories to inspire the further uptake of RDA outputs.
- **Submit your story here: https://www.rd-alliance.org/tell-your-rda-adoption-story**

CODATA Data Science Journal CfP

- **RDA special collection themes:**
  - Results produced by an IG or WG;
  - Description of an Adoption Case outlining how a specific recommendation or output has been implemented;
  - Other types of work related to RDA activities.
- RDA Europe 4.0 still has funds available for the publication of articles in DSJ
- Open to all interested applicants regardless of their geographical provenance.
- **Deadline 17 July**

Thursday 18th June

12:00 UTC

Disseminate, Link & Find

An increasing number of publishers and journals are implementing policies that require or recommend published articles to be accompanied by the underlying research data.

1. **Data Discovery Paradigms IG**
   - **Survey on the practices in data search services**
     Mingfang Wu (ARDC)
   - **Eleven quick tips and User requirements and recommendations**
     Fotis Psomopoulos (INAB CERTH)

Followed by Q&A

2. **FAIR data maturity model: specification and guidelines**

   Shelley Stall (AGU)

Followed by Q&A

3. **FAIR data maturity model: specification and guidelines** - FAIRsFAIR Adoption story

   Patricia Herterich (DCC)

Followed by Q&A

Data Discovery Paradigms IG

Relevancy Ranking Task Force

RDA Global Adoption week
18 June 2020

research data sharing without barriers
rd-alliance.org

Search for Data

Q Search

System view

Crawling/ Aggregating

Parsing

Indexing

Matching

Ranking

Query correction, recommendation

Snippet generation

data.gov

earthdata.nasa .gov

World Wide Science

Research data Australia

DataCite

Dryad

Zenodo

research data sharing without barriers
rd-alliance.org

RDA
RESEARCH DATA ALLIANCE

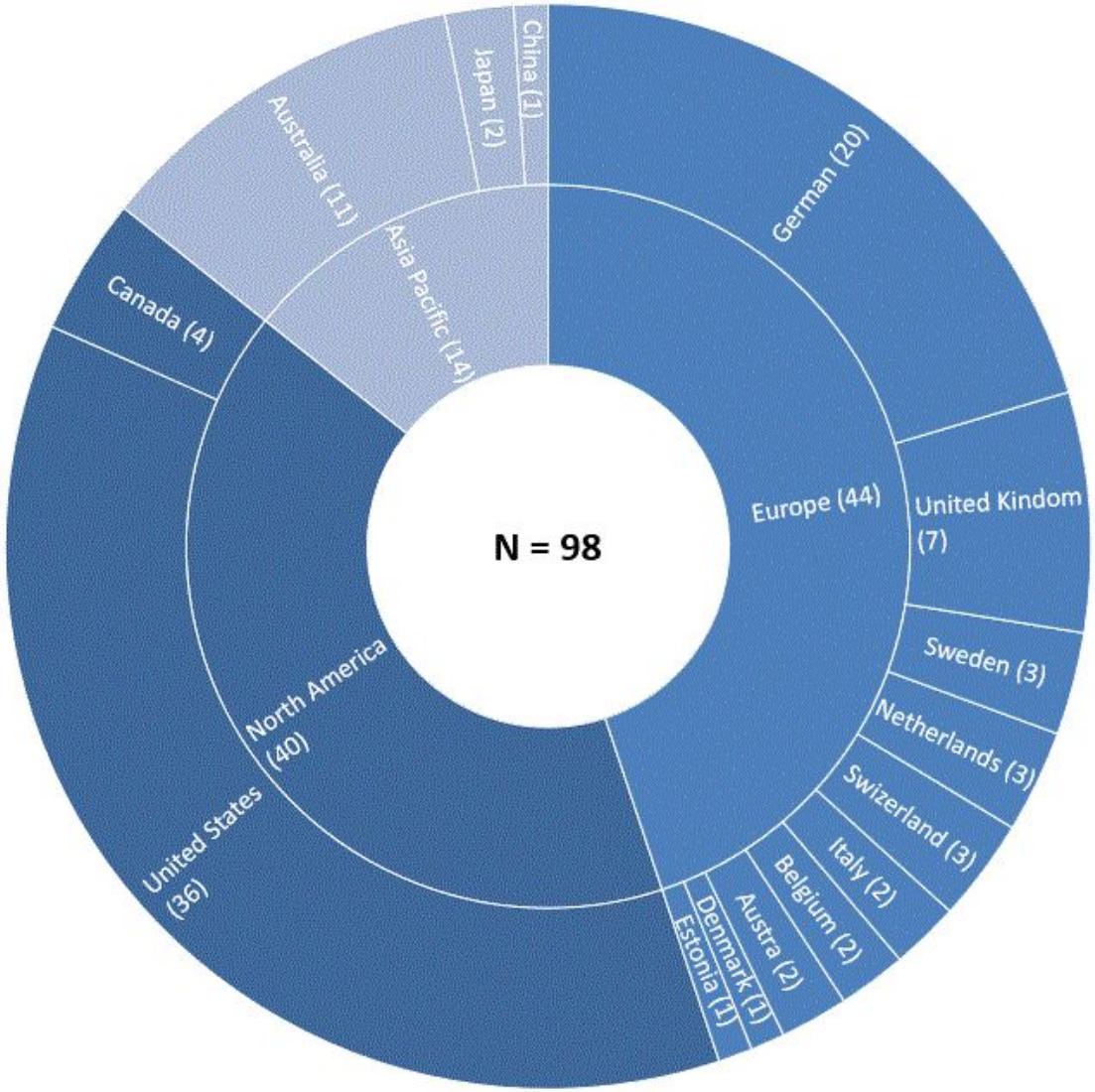
Survey goals

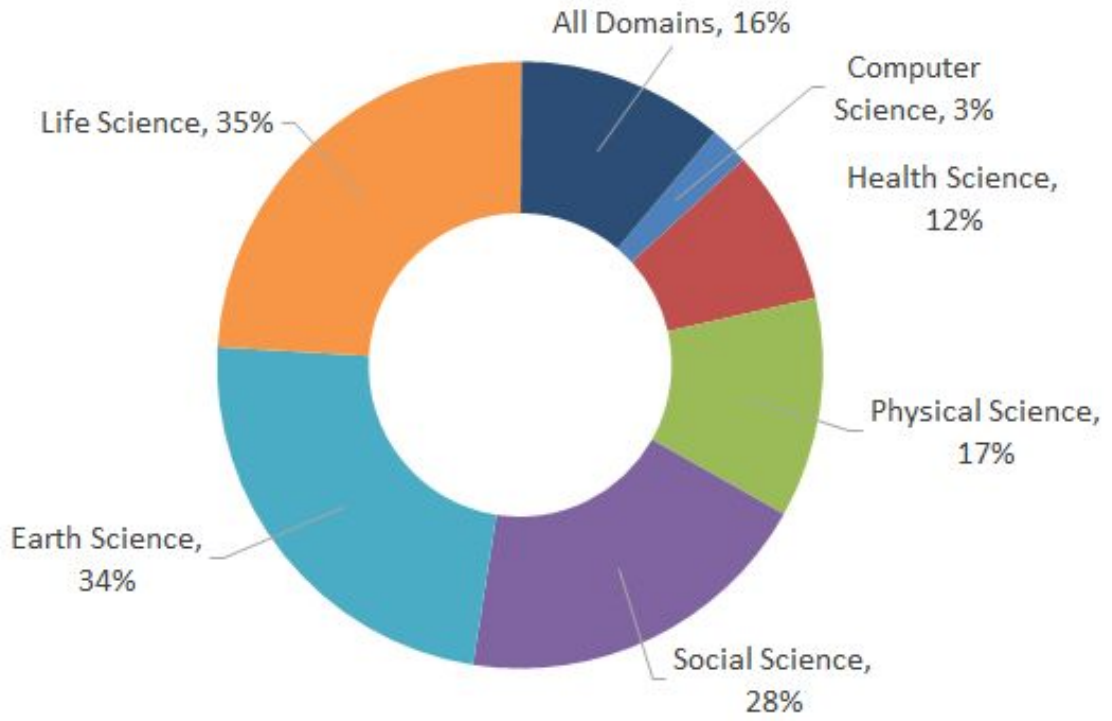
- Investigate what data search systems and ranking models have been deployed.
- Serve as a benchmark to be looked back on in future to assess how much and in what ways data search has improved.
- Identify potential collaborative projects from the Survey

Survey Design (33 Questions)

1. What are characteristics of each repositories? (5)
2. What are system configurations (e.g., ranking model, index methods, query methods)? (7)
3. What are evaluation methods and benchmark? (10)
4. What methods have been used to boost search-ability to web search engines? (2)
5. What other technologies or system configurations have been employed? (5)
6. Wish list for future activities for the RDA relevance task force (2)

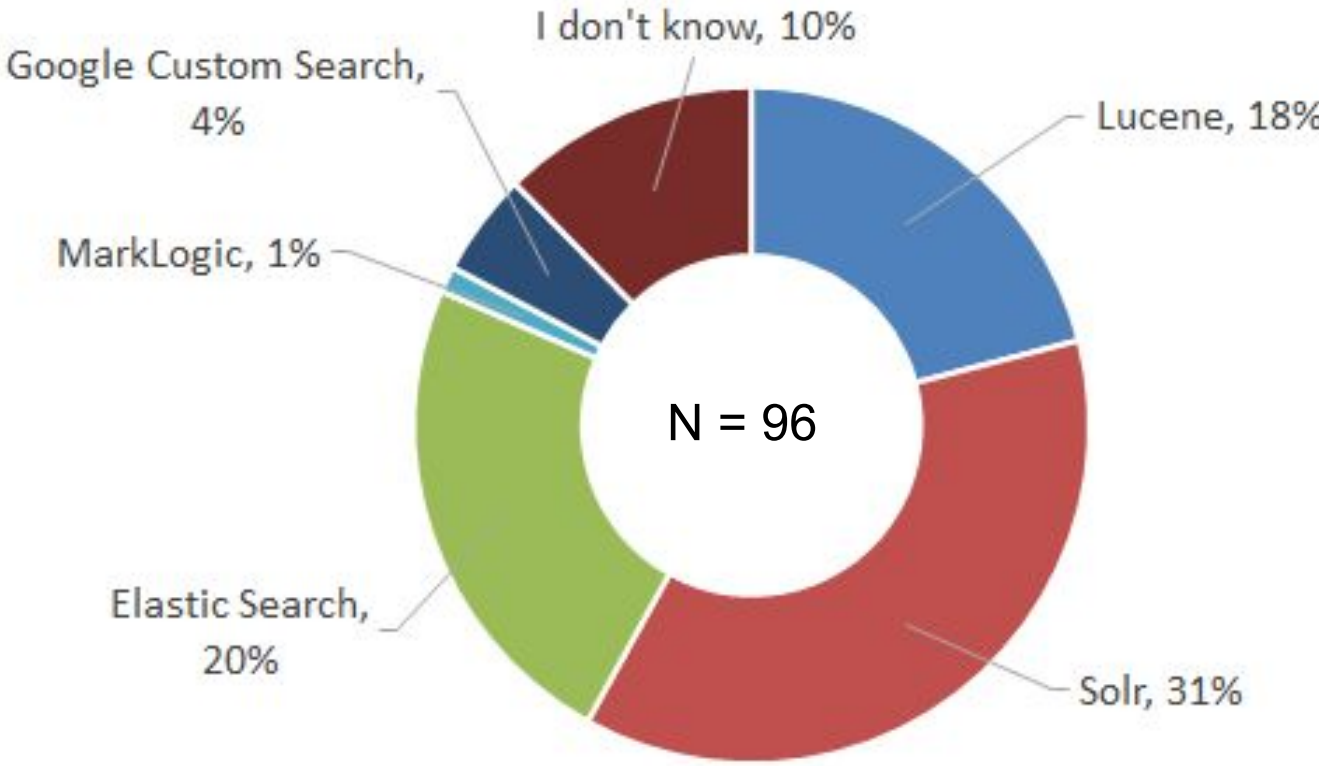
Participants background

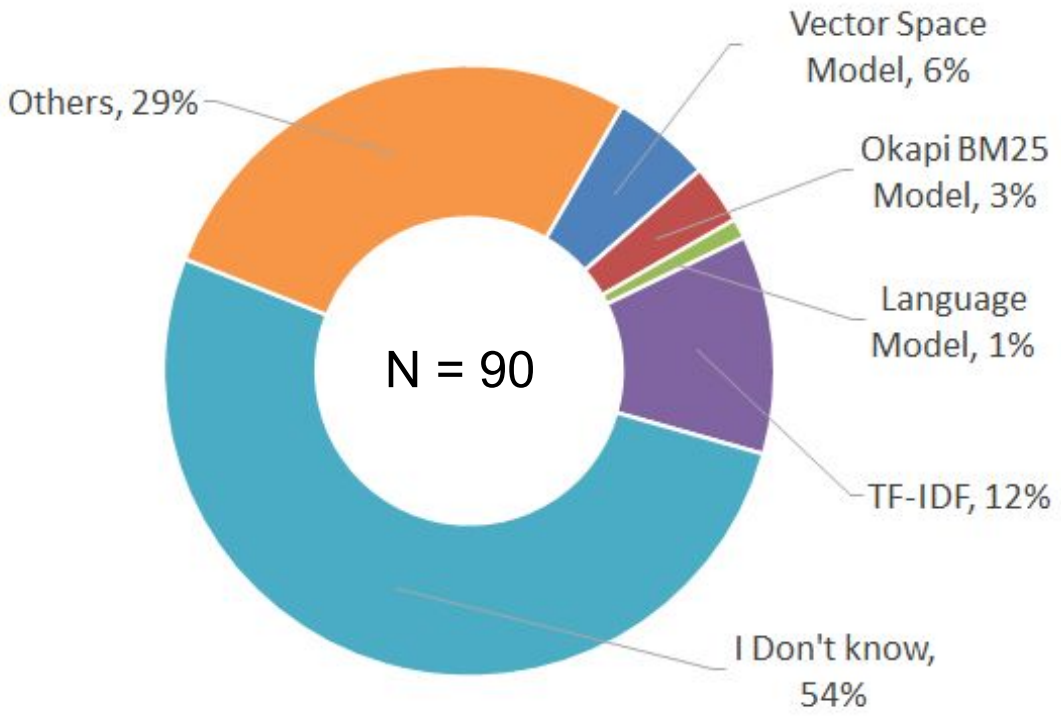




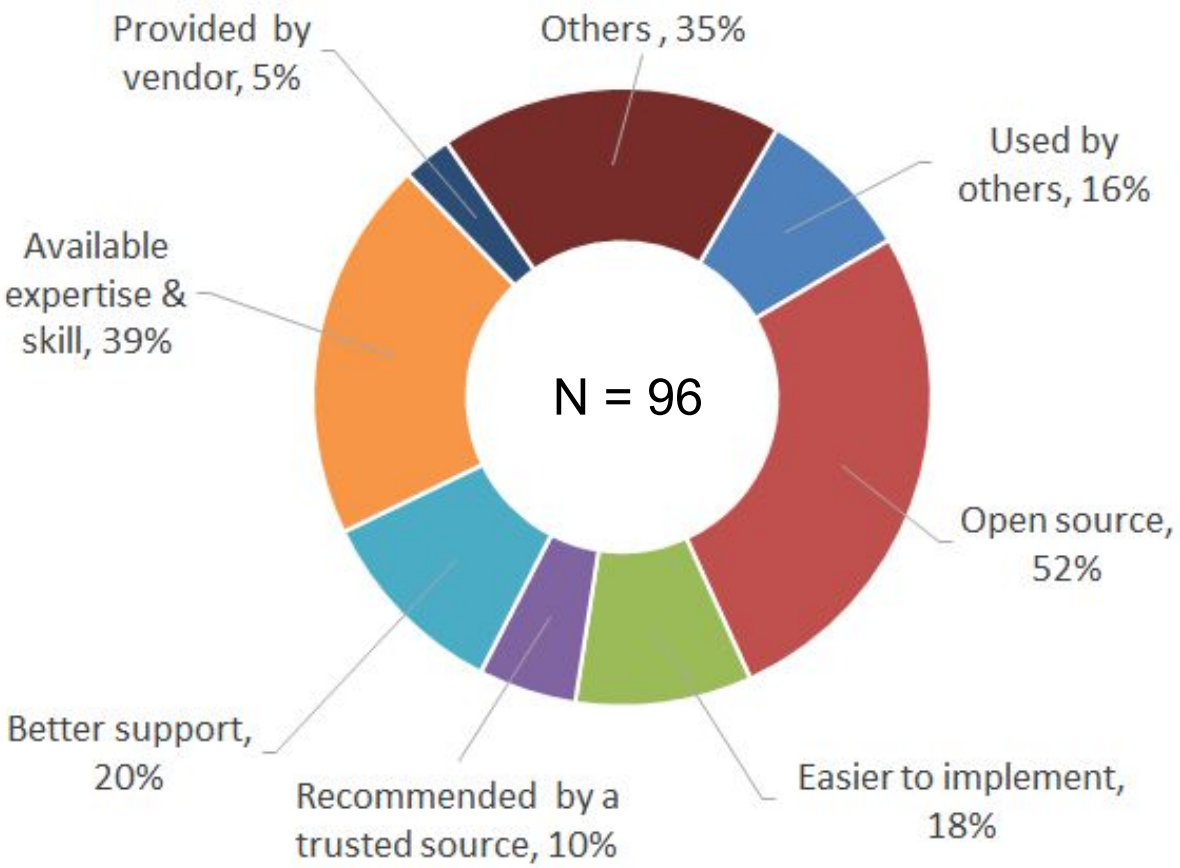
Survey result highlights ...

Data repositories use common search systems





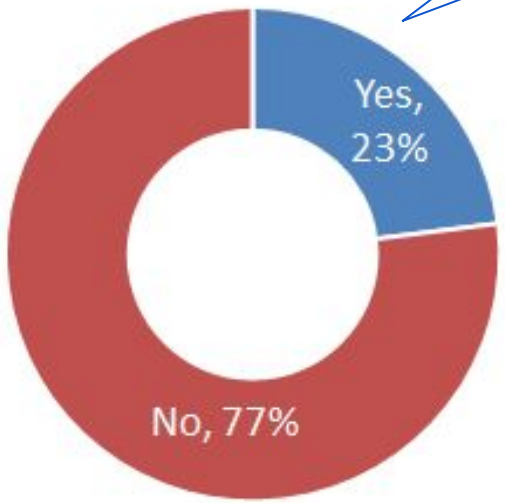
Open source and available skills are top reasons for choosing a search system

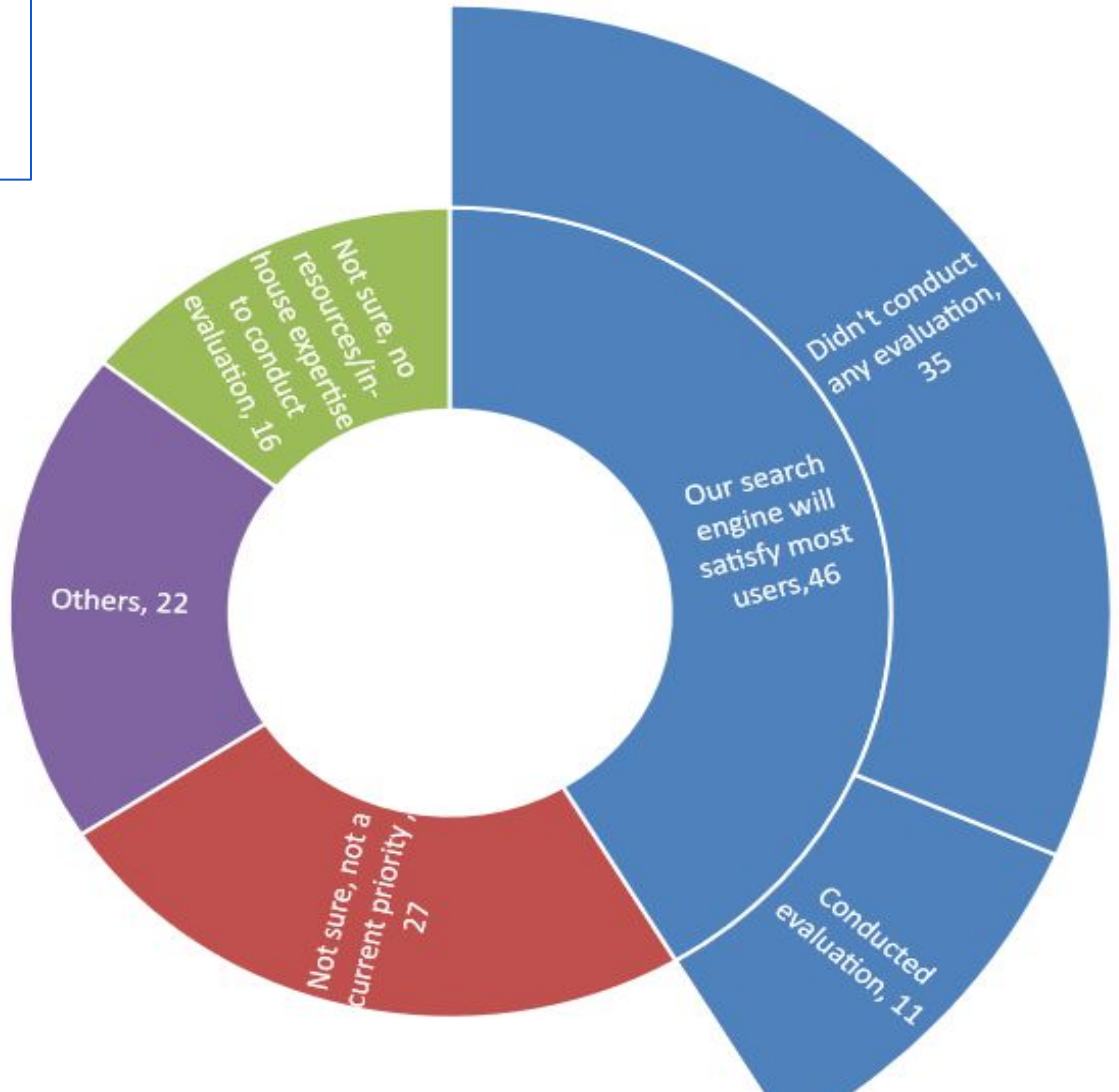


Majority didn't conduct any kind of evaluations

9 Created test collection
11 Informal evaluation
6 Log analysis

No performance measure was provided





Summary

- Repositories desire guidelines for improving relevancy ranking in their data search system, with small repositories having the greatest need.
- Repositories understand that their search systems need to be evaluated and improved, but often lack the resources (time and/or expertise) to explore and evaluate the available options.
- The study concludes that there is an opportunity for people working in the search space to collaborate, to build test collections and other efforts that offer the greatest improvements in search services at the lowest cost.

Khalsa, SiriJodha; Cotroneo, Peter; Wu, Mingfang (2018), “A survey of current practices in data search services”, Mendeley Data, v1 http://dx.doi.org/10.17632/7j43z6n22z.1

Thank you

Contact:
mingfang.wu@ardc.edu.au
sjsk@nsidc.org
fpsom@certh.gr

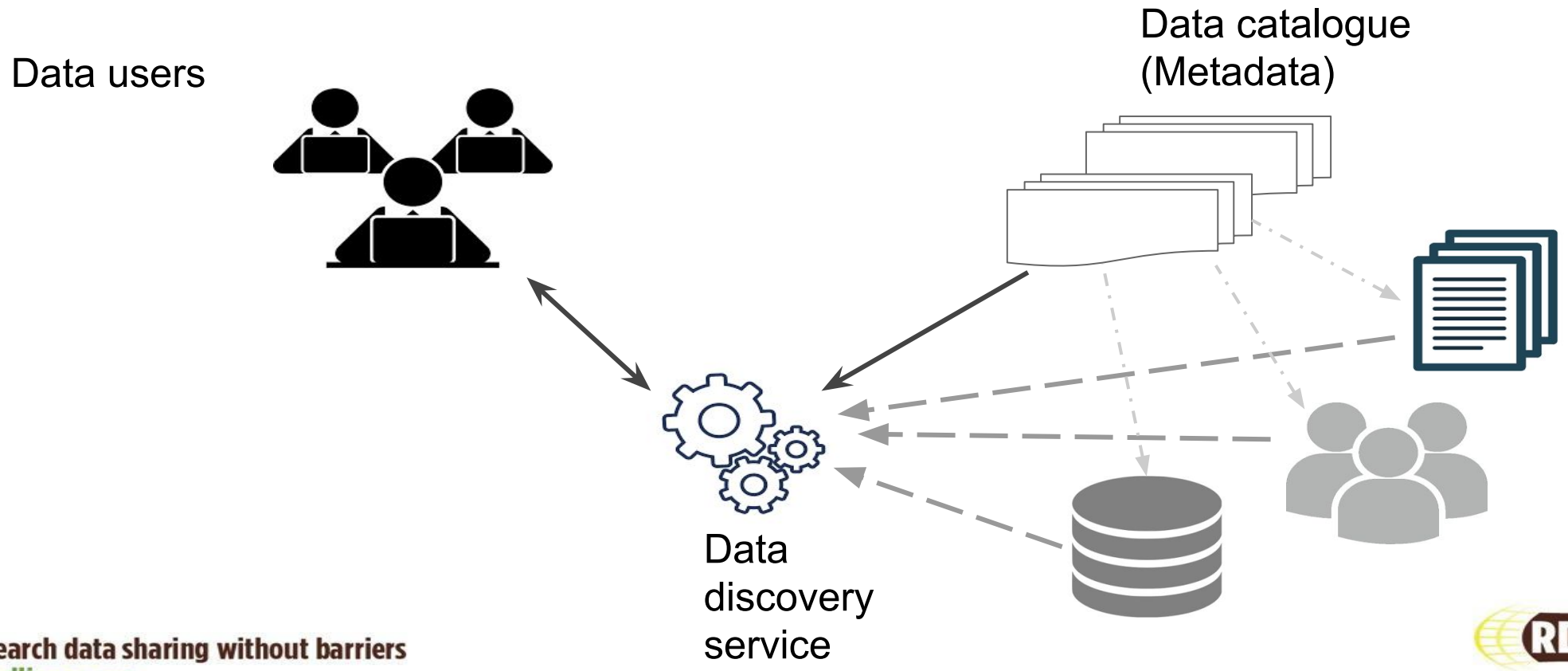
Data Discovery Paradigms Interest Group

RDA Global Adoption week
18 June 2020

research data sharing without barriers
rd-alliance.org

DDP Interest Group: Motivation

Helping to make research data **Findable** to support users in discovering data.



DDP Interest Group: Objective

- Provide a forum where representatives across the spectrum of stakeholders and roles can explore how to improve data discovery.
- Produce actionable recommendations for data producers, data repositories, data services providers and data seekers.

Output I - Eleven quick tips for finding research data

Tip 1: Think about the data you need and why you need them.
Tip 2: Select the most appropriate resource.
Tip 3: Construct your query strategically.
Tip 4: Make the repository work for you.
Tip 5: Refine your search.
Tip 6: Assess data relevance and fitness-for-use.
Tip 7: Save your search and data- source details.
Tip 8: Look for data services, not just data.
Tip 9: Monitor the latest data.
Tip 10: Treat sensitive data responsibly.
Tip 11: Give back (cite and share data).

Best practices for data seeker

Can be used for learning and research skills training

Gregory K, Khalsa SJ, Michener WK, Psomopoulos FE, de Waard A, Wu M (2018) Eleven quick tips for finding research data. PLoS Comput Biol 14(4): e1006038. https://doi.org/10.1371/journal.pcbi.1006038

(8124 views, 2345 downloads)

Output 2 - User Requirements for a data repository

Nine requirements (from 79 use cases)

- Indication of data availability
- Connection of data with person/institution/paper/citations/grants
- Fully annotated data
- Filtering of data based on specific criteria on multiple fields at the same time
- Cross-referencing of data
- Visual analytics/inspections of data/thumbnail preview
- Sharing data in a collaborative environment
- Accompanying educational/training material
- Portal functionality similar to other established academic portals

Data repository operators can use the requirements for the following purposes:

- As a checklist for designing and implementing a data service portal.
- For existing data discovery services, the list of requirements can be used as guidelines for heuristic evaluation of a specific data discovery service, and therefore plan for future improvements when necessary.
- In the era of big data, research on data discovery paradigms is at an all-time high. A user's perspective provides a strong foundation on which to construct the paradigms of the future.

Output 2 - Recommendations for Data Repositories to make data discovery

Recommendations:

- Multiple query interfaces
- Multiple access points
- Assessable search result
- Readable and analysable metadata records
- Available bibliographic references
- Available data usage statistics
- Consistent interface
- Identifiable duplicats
- Findable from web search engines
- Interoperability with other repositories

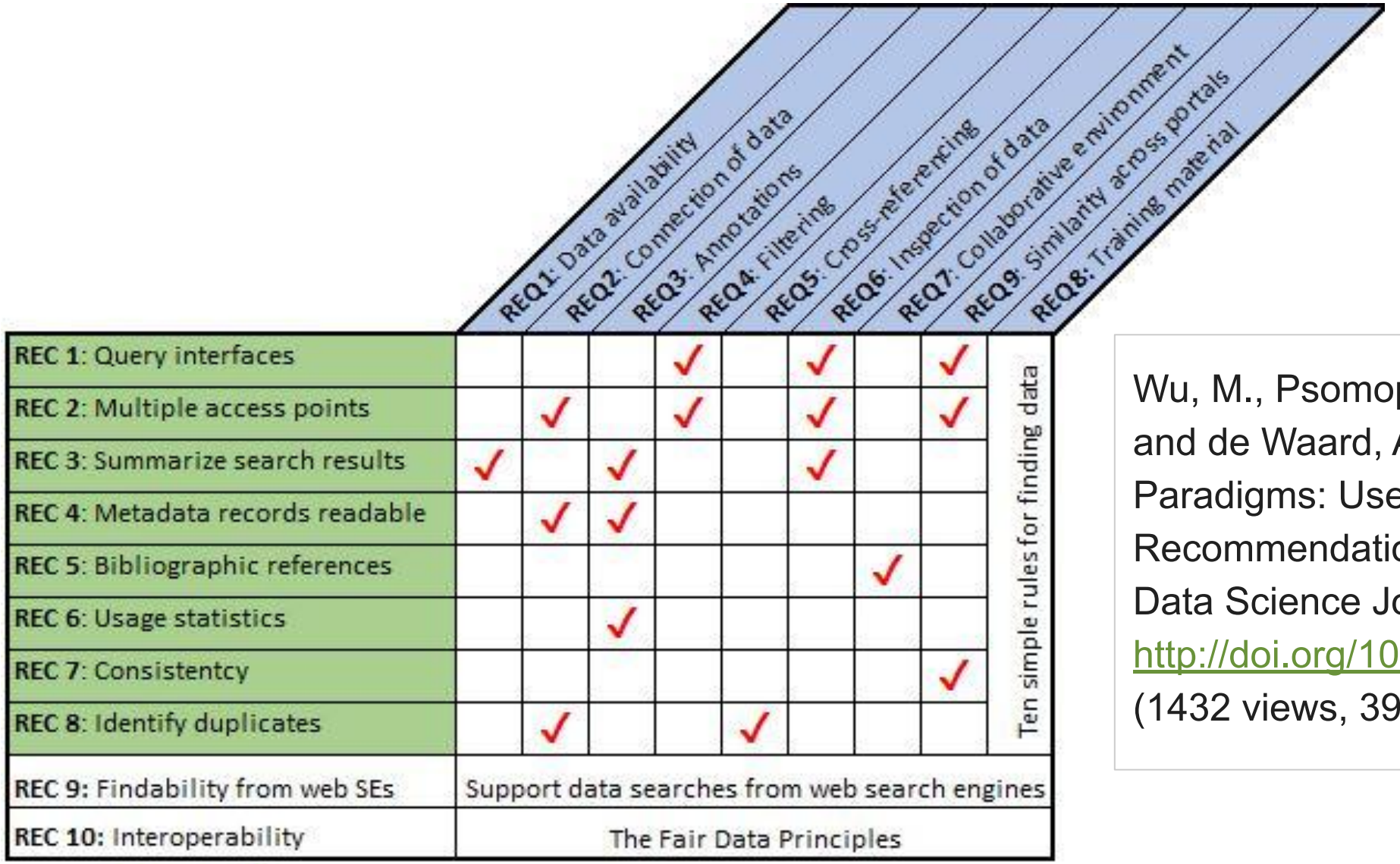
Data repositories can take the ten recommendations:

- As guidelines when implementing a new repository
- As a checklist when conducting heuristic evaluation of an existing repository.

Data repositories can implement all or prioritise their implementation based on their user needs and available resources.

Use cases published to Zenodo https://doi.org/10.5281/zenodo.1050976 (124 views, 73 downloads)

Output 2 - User Requirements and Recommendations for Data Repositories



Wu, M., Psomopoulos, F., Khalsa, S.J. and de Waard, A., 2019. Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories. Data Science Journal, 18(1), DOI: http://doi.org/10.5334/dsj-2019-003 (1432 views, 396 downloads

RDA RESEARCH DATA ALLIANCE

Contact

fpsom@certh.gr

mingfang.wu@ardc.edu.au

sjsk@nsidc.org

https://www.rd-alliance.org/groups/data-discovery-paradigms-ig

Adoption of the FAIR Data Maturity Model

18 June 2020

Context

The principles are **NOT** strict

- **Ambiguity**
- Wide range of **interpretations** of FAIRness

Different **FAIR Assessment** Frameworks

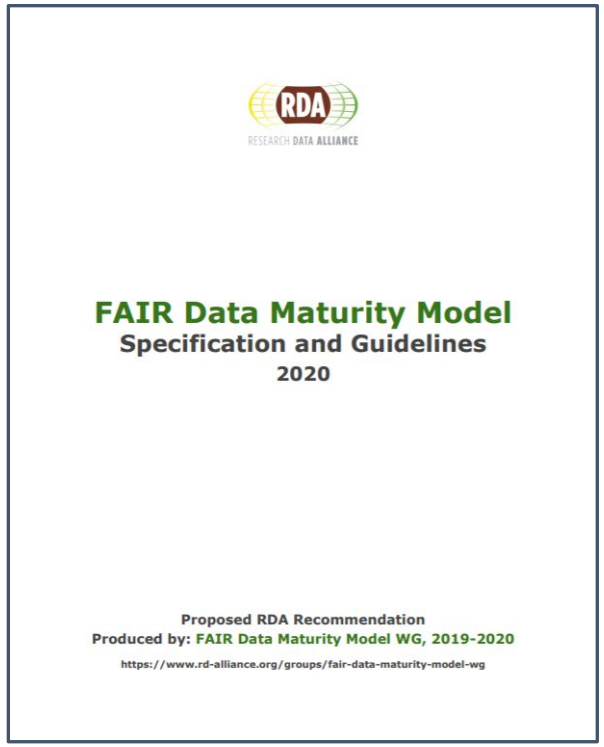
- Different metrics
- No comparison of results
- No benchmark

SOLUTION is to bring together **stakeholders** to build on **existing approaches** and **expertise**

- Set of **core assessment criteria** for FAIRness
- FAIR **data maturity model & toolset**
- FAIR data **checklist**
- RDA recommendation

Join the **RDA** Working Group: RDA WG web page | GitHub

Public review period complete now to council



THANKS TO ALL REVIEWERS

3600+ page views

14 comments

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines

Adoption examples

Early adopters – Experience sharing

- **Ge Peng | NOAA**
- **Anusuriya Devaraju | FAIRsFAIR**

… will share their relevant experience with regard to the adoption of the FDMM and answer to the following questions;

1. **What is the level of adoption at your organisation? (E.g., pilot, production, ...)**
2. **Do you plan to continue to use the Recommendation?**
3. **Did you need to modify the Recommendation for your use?**
4. **Can you give an estimate of how much time / effort you have spent on the adoption so far?**
5. **What's your overall experience? (E.g., Very Good, Good, Fair, Poor)**
6. **Would you do it again?**

Evaluating the FAIRness of Environmental Data

– Application of the RDA FAIR Data Maturity Indicators

Ge Peng, PhD

Cooperative Institute for Satellite Earth System Studies (CISESS) Between
U.S. National Oceanic and Atmospheric Administration (NOAA) and North Carolina State University
at NOAA National Centers for Environmental Information (NCEI)

#9 Workshop of the RDA FAIR Data Maturity Model Working Group, May 20–21, 2020

Purposes of Pilot Application

- Examine the relevancy of the RDA FAIR DMIs (v0.04)
- **Baseline the FAIRness of NCEI managed data**
  - In particular, *OneStop*-Ready datasets,
    - *OneStop* project was Initiated in 2015 to improve discovery and access services for NOAA datasets.
  - What worked?
- **Identify potential gaps & define path forward in NCEI data sharing practices**

Adopting OAIS RM & DSMM Helped!

Mapping FAIR Data Principles to NCEI/CICS-NC Data Stewardship Maturity Matrix (DSMM)

FAIR Data Principles (Wilkinson et al. 2016)	DSMM Key Components								
	Preservability	Accessibility	Usability	Production Sustainability	Data Quality Assurance	Data Quality Control/Monitoring	Data Quality Assessment	Transparency /Traceability	Data Integrity
F1. (meta)data are assigned a globally unique and eternally persistent identifier								L3	
F2. data are described with rich metadata (defined by R1 below)	L3		L3					L5	
F3. metadata clearly and explicitly include the identifier of the data it describes	L3		L3					L3	
F4. (meta)data are registered or indexed in a searchable resource		L2 & L3							
A1. (meta)data are retrievable by their identifier using a standardised communications protocol		L2 & L3	L3					L3	
A1.1. the protocol is open, free, and universally implementable		L3							
A1.2. the protocol allows for an authentication and authorization procedure, where necessary		L3							
A2. metadata are accessible, even when the data are no longer available		L2							
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	L3		L3						
I2. (meta)data use vocabularies that follow FAIR principles		L4							
I3. (meta)data include qualified references to other (meta)data	L3		L3						
R1. meta(data) are richly described with a plurality of accurate and relevant attributes	L3		L3						
R1.1. (meta)data are released with a clear and accessible data usage licence	*		*						
R1.2. (meta)data are associated with detailed provenance									
R1.3. (meta)data meet domain-relevant community standards	L3		L3						

* Can be easily implemented via relevant metadata entity and modified document template

(Version: v00r01 20200403; POC: gpeng@ncsu.edu; CC-BY 4.0)

Many data stewardship quality attributes are not explicitly addressed by the FAIR Data Principles.

- Most of data are open by default,
- Use agreements or use constraints,
- CC license not yet explicitly included.

Path Forward

Improving the FAIRness of NCEI & NOAA Data

- **Explicitly include** a data usage license, e.g. CC-BY 4.0; CC0, in the metadata record:
  - Discussions are on-going,
  - Procedure is under development.

Extending the Application Scope – under discussion

- **Assess:** 200+ additional NCEI datasets,
  - produced by NCEI's Center for Weather and Climate, *various stages of OneStop-ready.*
- **Examine** the scalability of the evaluation.

Integrating Assessment Results - Fairly

- Community guidelines – consistently curating and representing dataset quality information,
- Virtual workshop on July 13, 2020 – bringing together international domain experts,
- Contact me at gpeng@ncsu.edu if interested in participating or contributing.

FAIRSFAIR
Fostering Fair Data Practices in Europe

RDA FAIR Data Maturity Model Adoption (Impression and Experience)

Anusuriya Devaraju & Hervé L'Hours
(on behalf of FAIRsFAIR)

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

Repository Certification

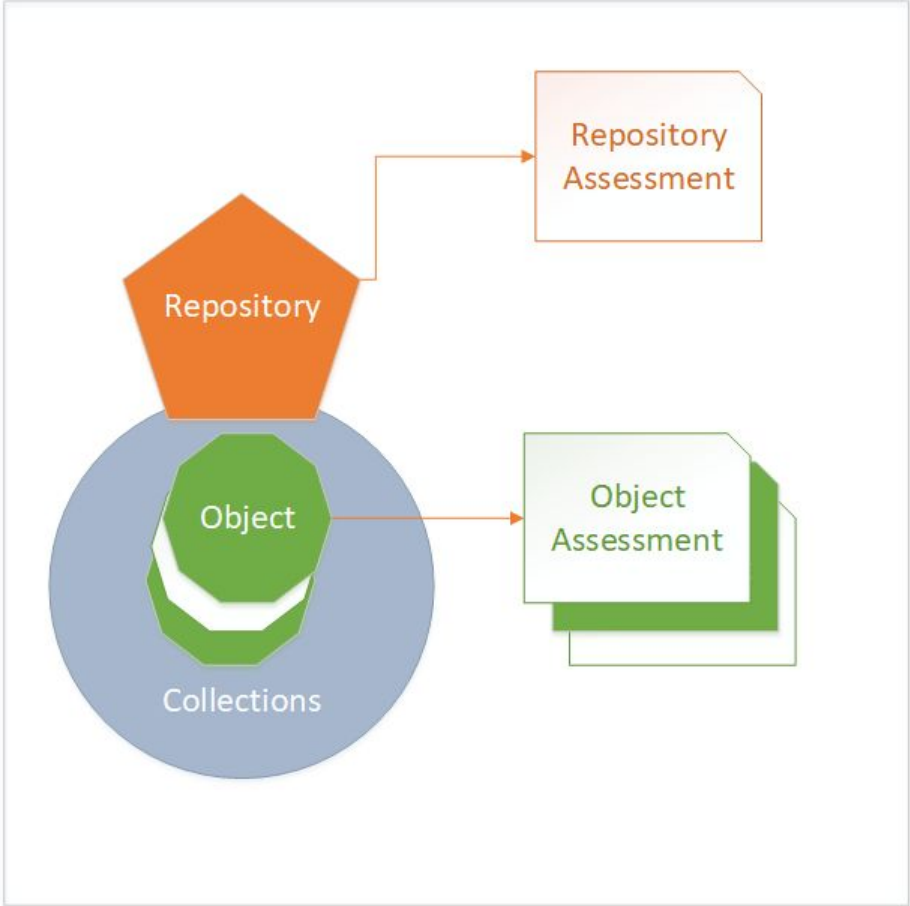
- CoreTrustSeal follows a self-assessment and peer review model
- FAIRsFAIR is offering support with a CoreTrustSeal+FAIR angle
- Map object characteristics to where repositories can enable FAIR

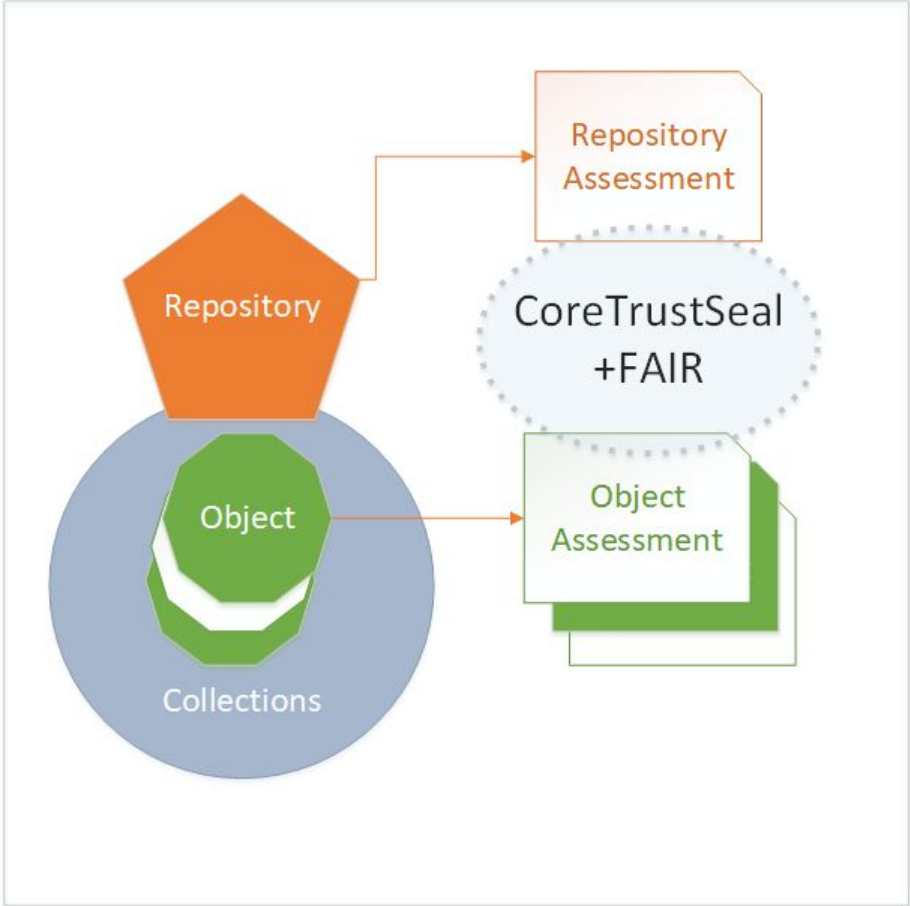
Repository Certification

- CoreTrustSeal follows a self-assessment and peer review model
- FAIRsFAIR is offering support with a CoreTrustSeal+FAIR angle
- Map object characteristics to where repositories can enable FAIR

Later:

- **Integrate object evaluation outcomes**


Repository
Assessment
Repository
Object
Object
Assessment
Collections


Repository
Repository Assessment
CoreTrustSeal +FAIR
Object
Object Assessment
Collections

Overall Adoption Experience

- The recommendation should be used as a starting reference point for data FAIRness assessment.
- Presentation - specification and guidelines are well structured!
- ‘What’ aspect of FAIR assessment
  - Descriptions of indicators are very helpful!
  - Suggestion - Include priority level next to each of the indicators.
  - Essential I-indicators missing (needs further work or not important?)
- ‘How’ aspect of FAIR assessment
  - Context matters (e.g., practices, data types)
  - Assessment details not always provide sufficient detail to implement tests.
  - Potential supporting technologies and services should be described.

Next steps

- Reach out to your communities as for the publishing of the **FAIR data maturity model: specification and guidelines** (i.e. RDA recommendation)
- Continuously provide feedback to the Editorial Team and pass on information with regards to the use of the **FAIR data maturity model: specification and guidelines** (i.e. RDA recommendation)

The editorial team will look into a release calendar and change management schedule

WORKSHOP #10

Possibly **September 2020**

Thank you!

FAIRSFAIR
Fostering Fair Data Practices in Europe

FAIR data maturity model - FAIRsFAIR Adoption story

Patricia Herterich
(on behalf of FAIRsFAIR)

RDA Global Adoption week: Thursday, 18th June 2020 - Disseminate, Link and Find

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_EU

Outline

- FAIRsFAIR project
- FAIRsFAIR's CoreTrustSeal certification support
- FAIR assessment of digital (data) objects

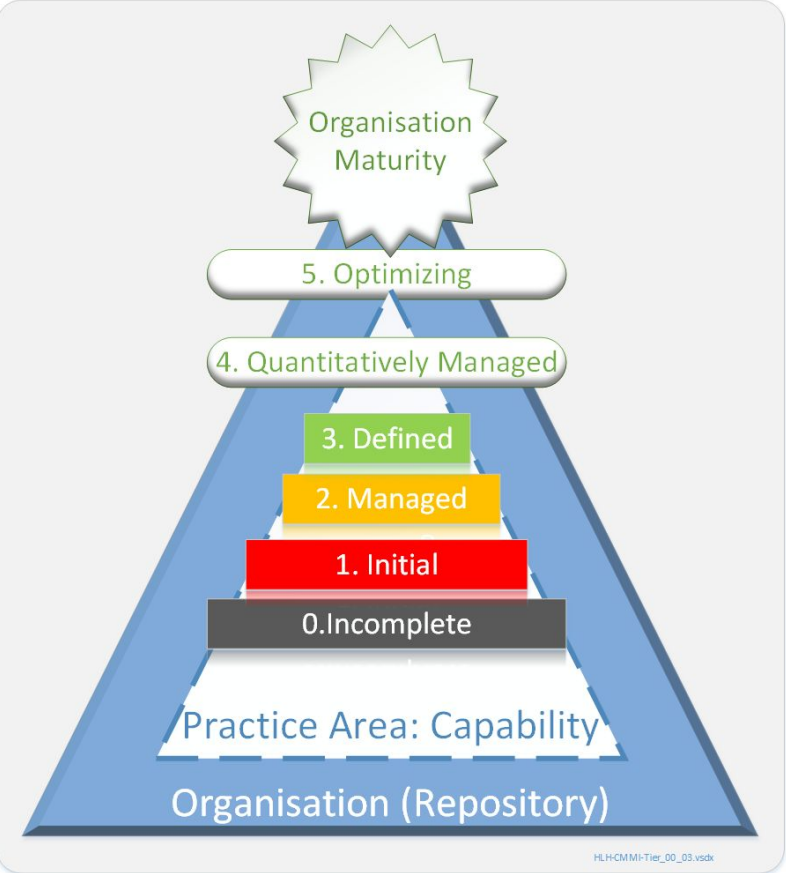
FAIRsFAIR - Fostering Fair Data Practices in Europe

https://www.fairsfair.eu

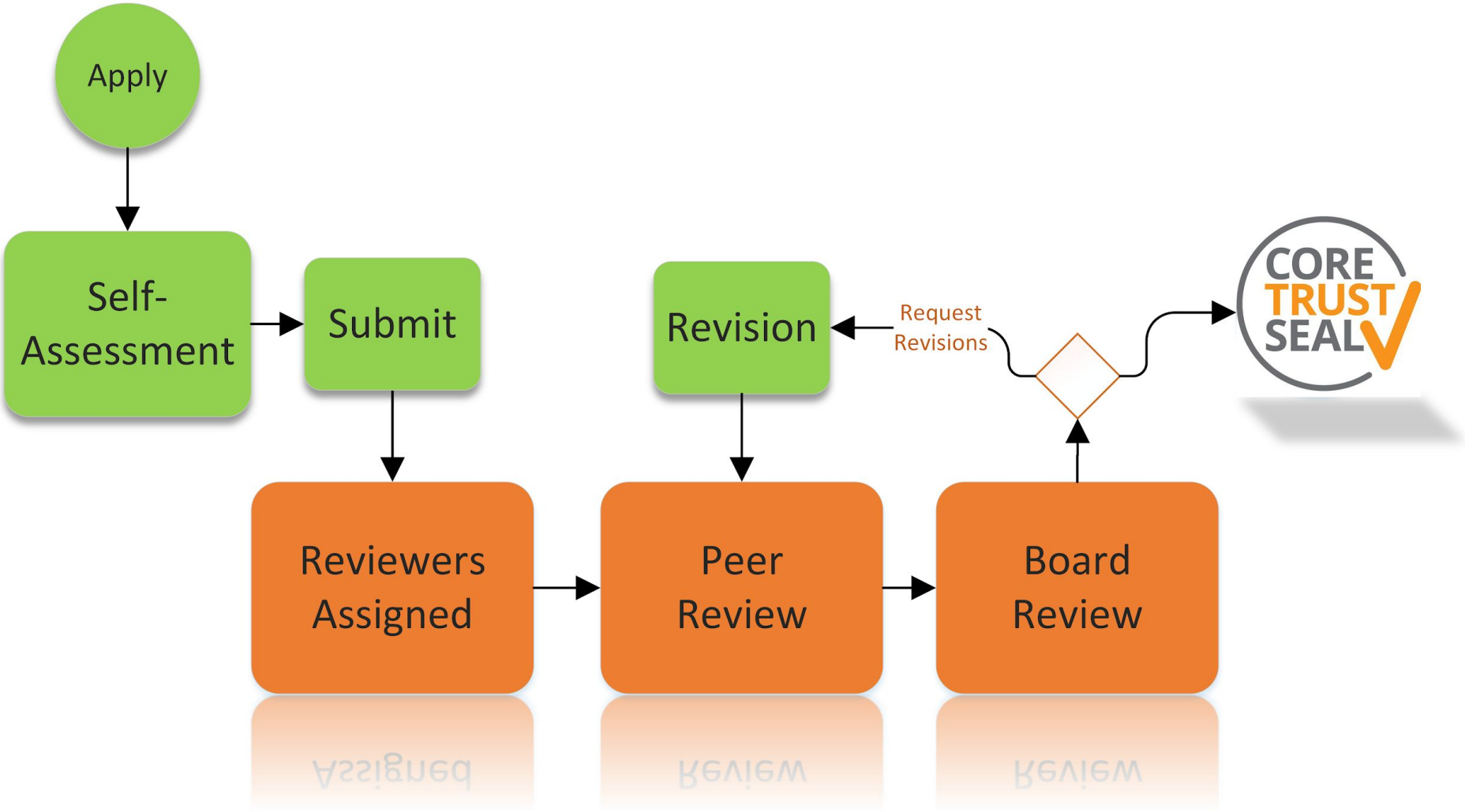
- Aims to supply practical solutions for the use of the FAIR principles throughout

FAIRsFAIR & RDA FAIR Data Maturity Model WG

- Involvement in RDA WG activities is mainly through WP4 (FAIR Certification):
  - Capability maturity models towards FAIR certification of repositories
  - FAIR assessment of digital (data) objects : Pilots (two primary use cases)

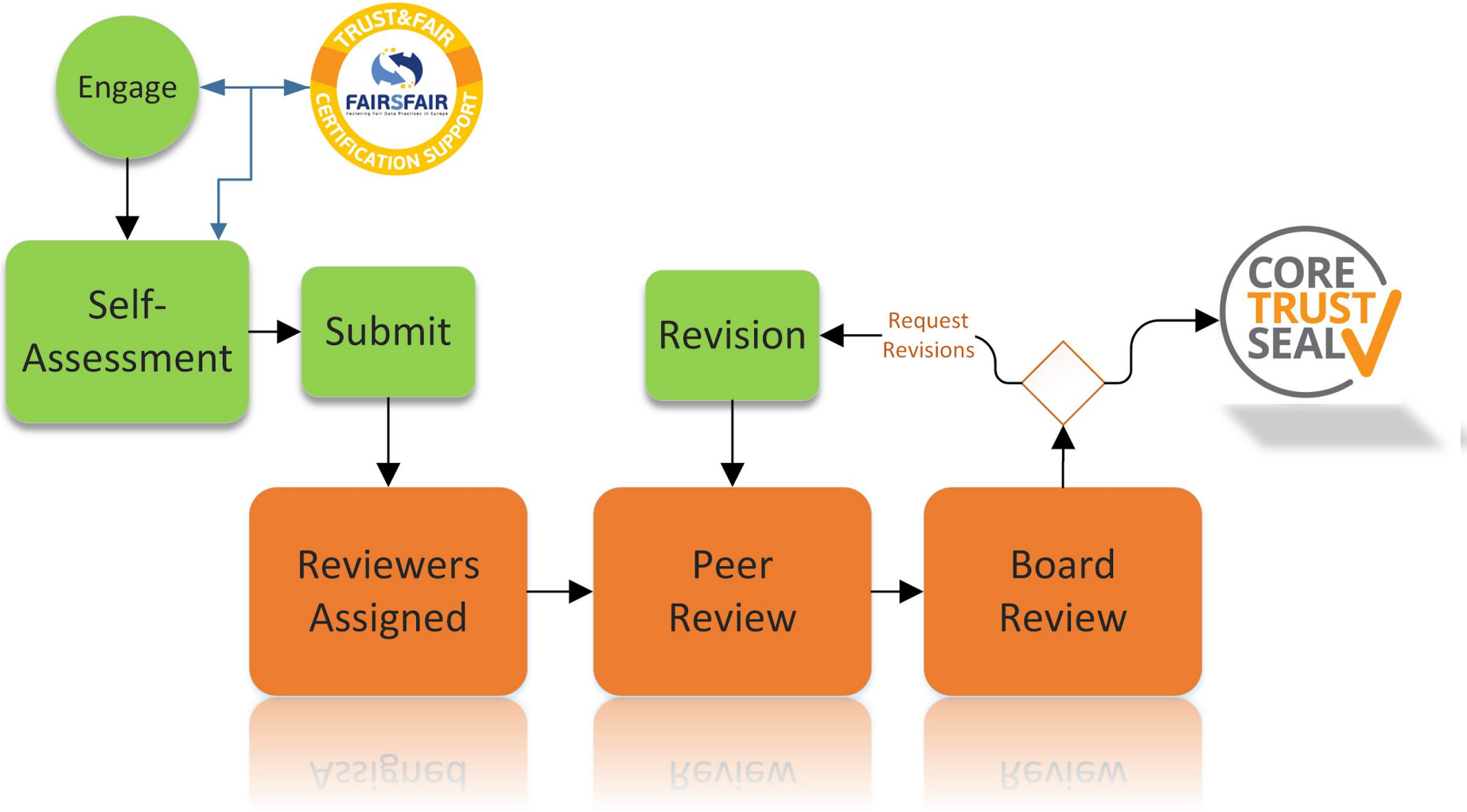


Repository Certification – Core Trust Seal process



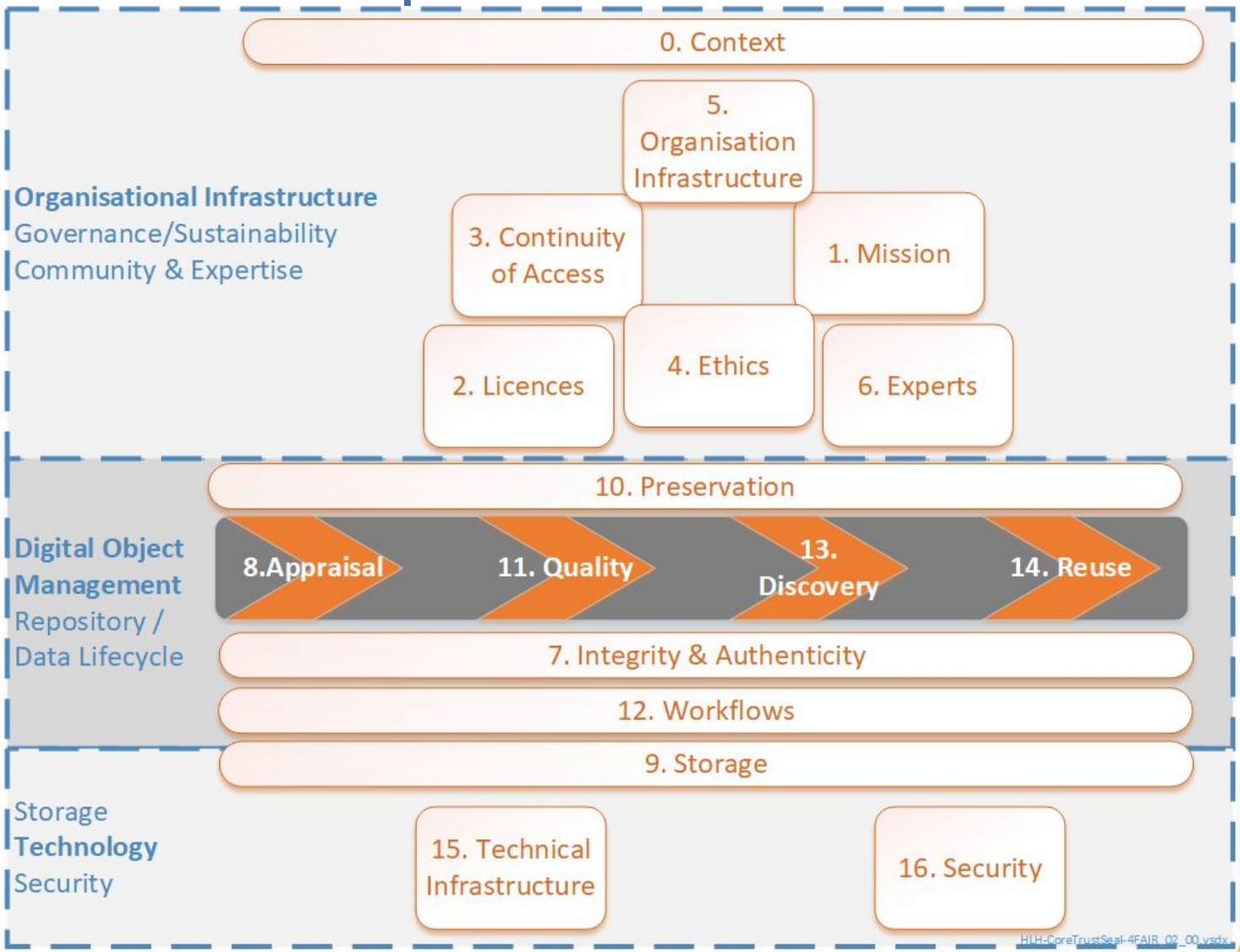
HLH-CoreTrustSeal-Process_00_03.vsdx

FAIRsFAIR CoreTrustSeal+FAIR certification support

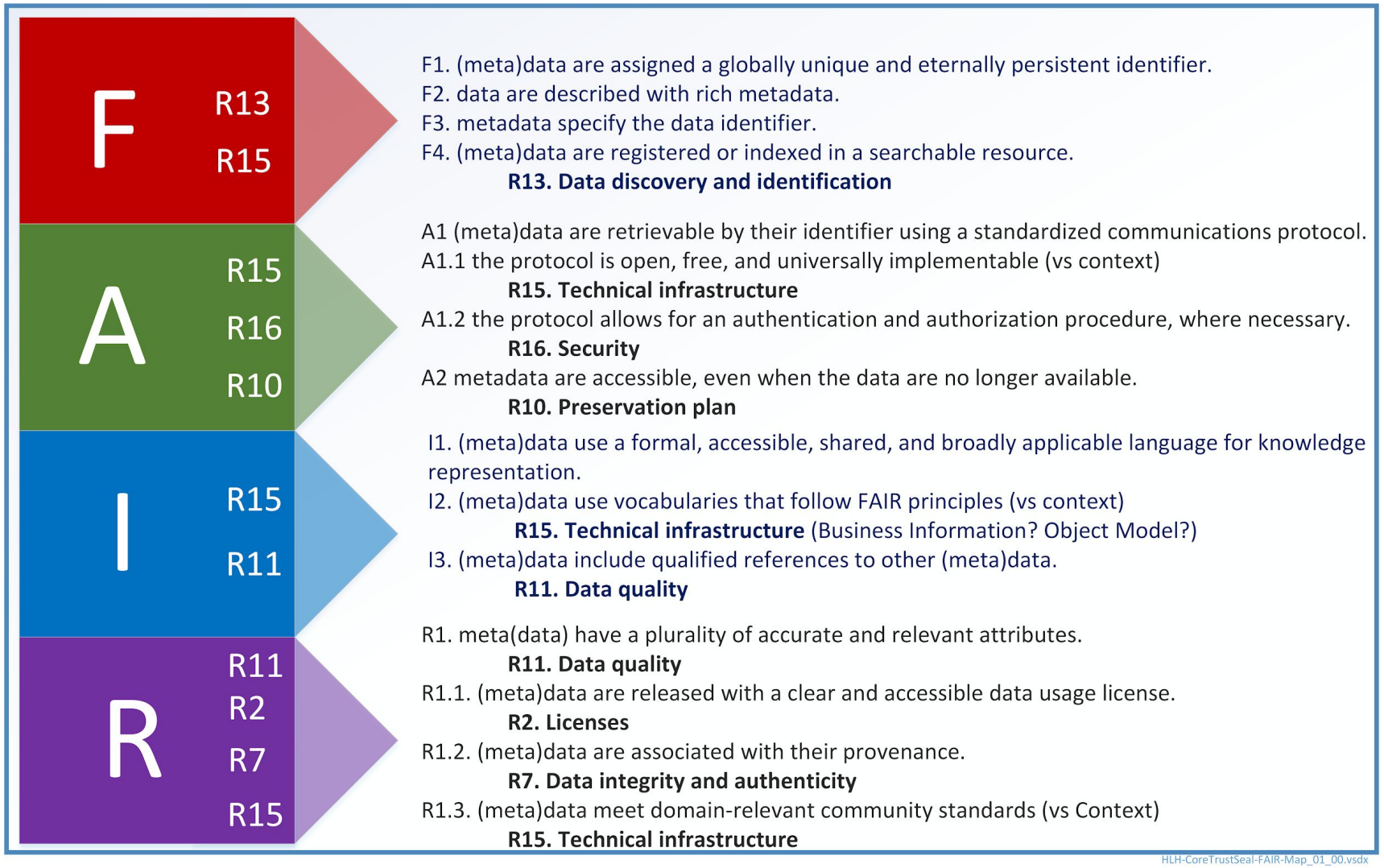


HLH-CoreTrustSeal-Process-FAIRsFAIR_00_03.vsdx

CoreTrustSeal requirements



Mapping CoreTrustSeal and FAIR



F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. metadata specify the data identifier.
F4. (meta)data are registered or indexed in a searchable resource.
R13. Data discovery and identification

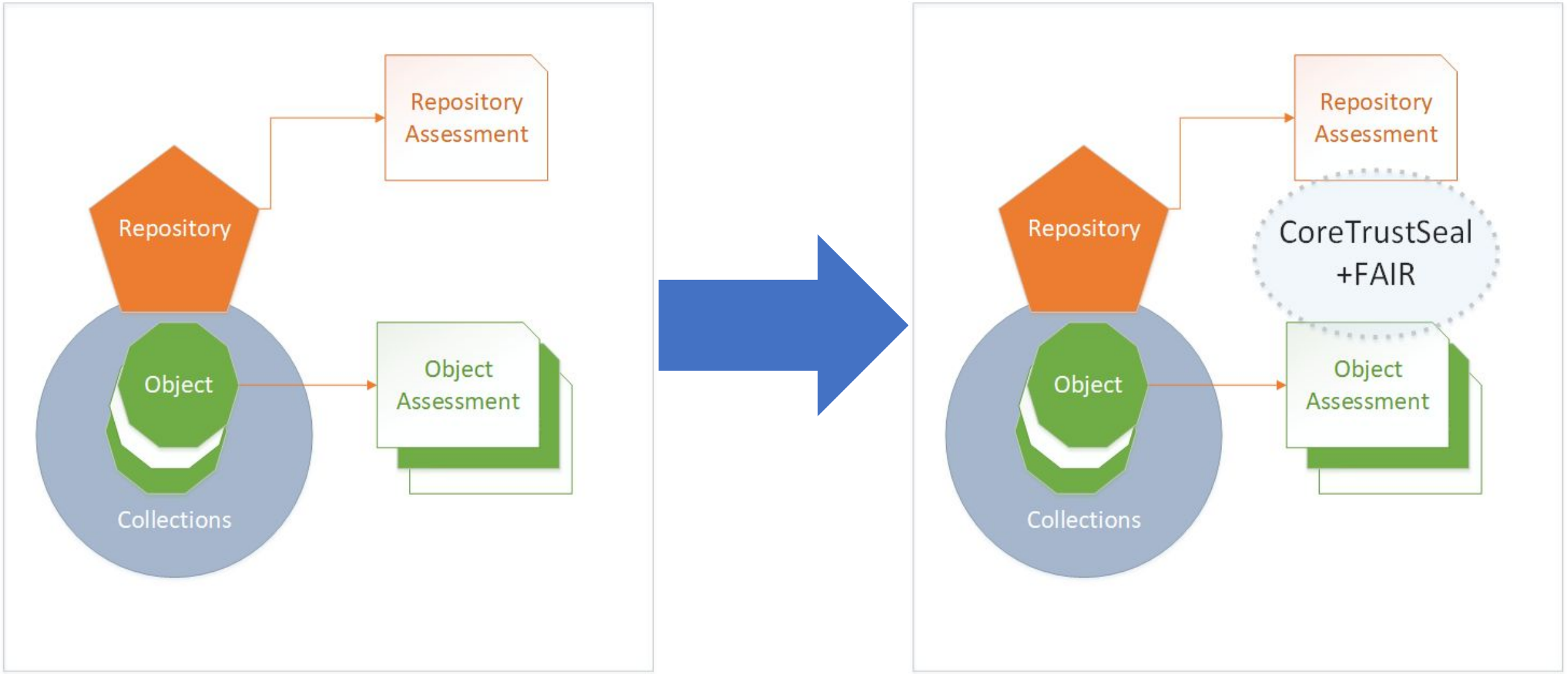
A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable (vs context)
R15. Technical infrastructure
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
R16. Security
A2 metadata are accessible, even when the data are no longer available.
R10. Preservation plan

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles (vs context)
R15. Technical infrastructure (Business Information? Object Model?)
I3. (meta)data include qualified references to other (meta)data.
R11. Data quality

R1. meta(data) have a plurality of accurate and relevant attributes.
R11. Data quality
R1.1. (meta)data are released with a clear and accessible data usage license.
R2. Licenses
R1.2. (meta)data are associated with their provenance.
R7. Data integrity and authenticity
R1.3. (meta)data meet domain-relevant community standards (vs Context)
R15. Technical infrastructure

HLH-CoreTrustSeal-FAIR-Map_01_00.vsdx

Unifying repository and object assessment



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'

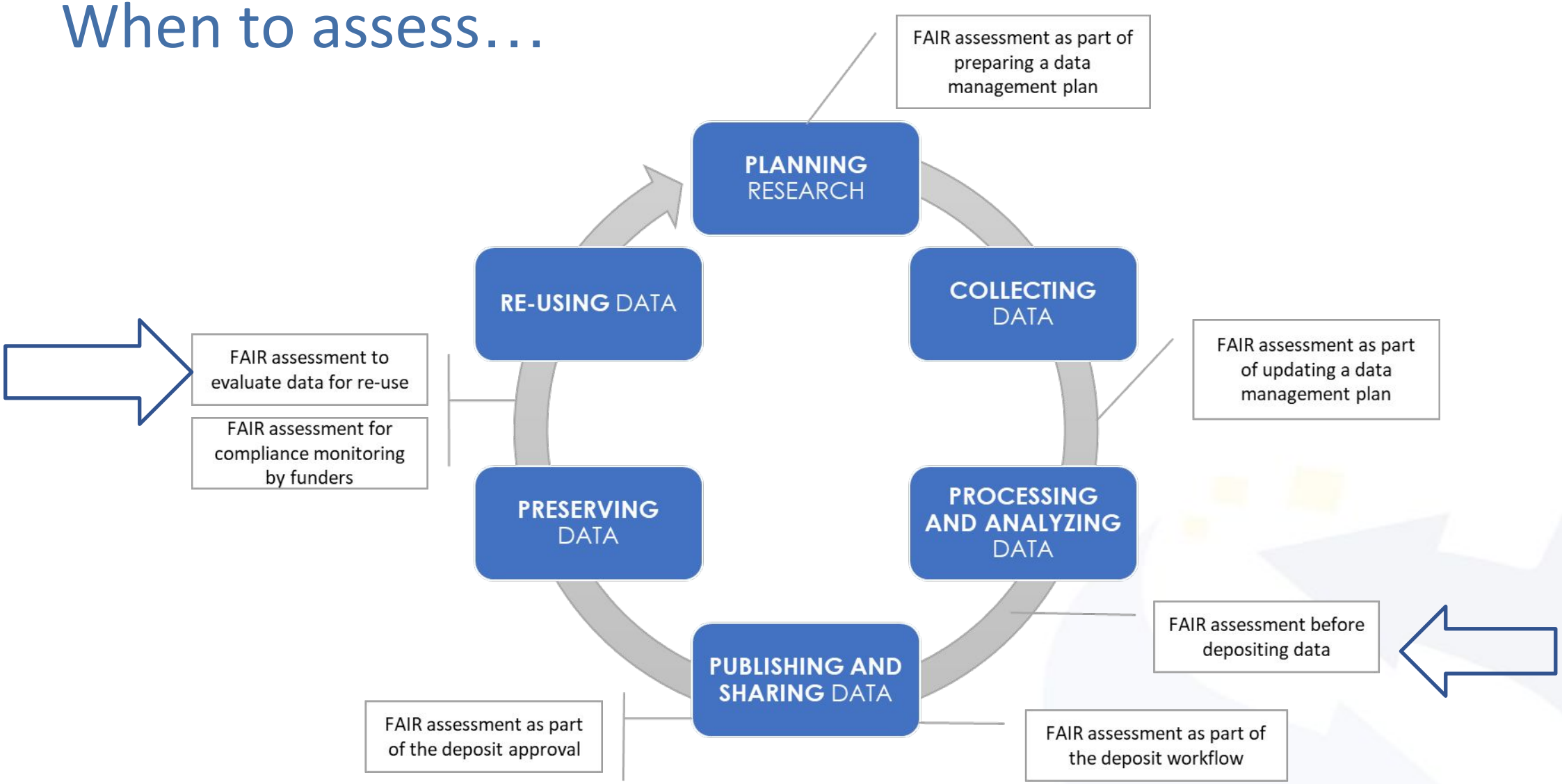
Research Data Alliance FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model: specification and guidelines - draft. Research Data Alliance. https://doi.org/10.15497/RDA00045

What? - The FAIRsFAIR Data Assessment Metrics

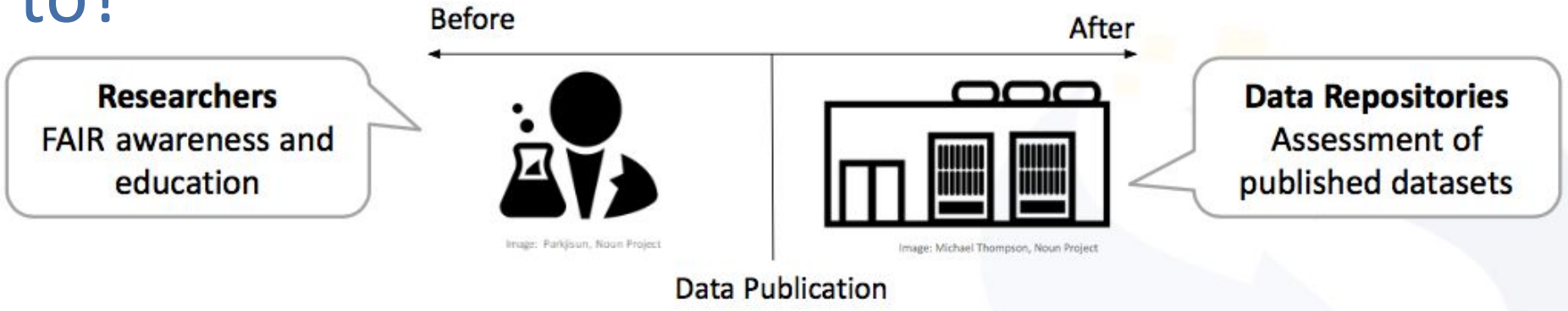
- There are 15 metrics built on existing work.
  - RDA FAIR Data Maturity Model
  - DANS Fairdat/FAIREnough
  - WDS/RDA Assessment of Data Fitness
- Iteratively improve and extend the metrics through a number of pilot tests.
- v0.3 will be released in August 2020

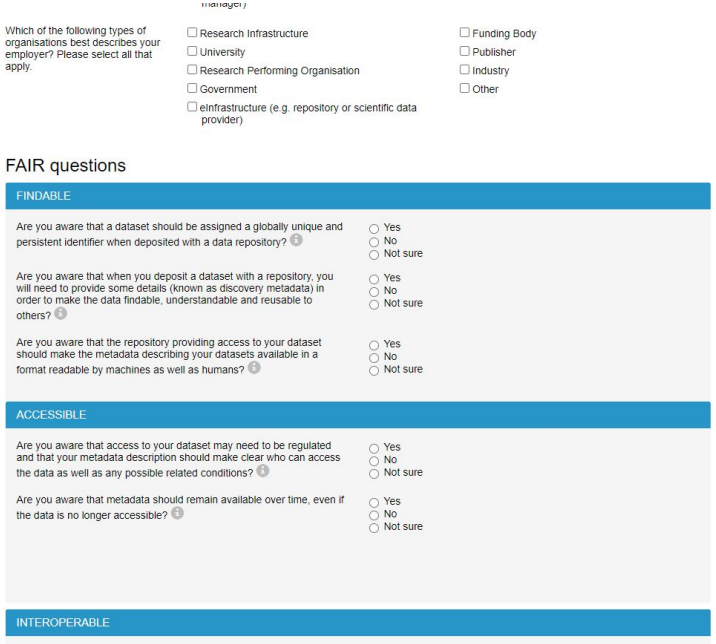
FIELD	DESCRIPTION
Metric Identifier	FsF-F1-01D
Metric Name	**Data** is assigned with a globally unique identifier.
Description	A data object may be assigned with a globally unique identifier such that it can be referenced unambiguously by humans or machines. Globally unique means an identifier should 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:	

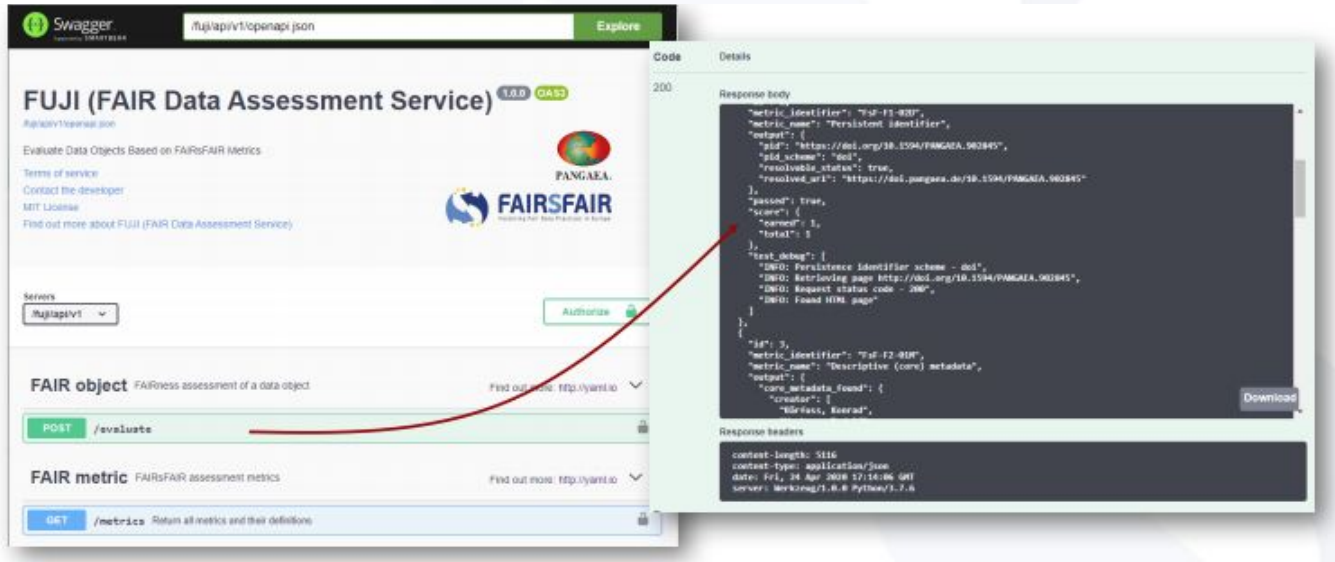
When to assess…



How to?







Toolset snippets: manual self-assessment awareness tool (left) and automated assessment tool (right)
Developments are available at https://github.com/FAIRsFAIR/WP4_FAIRsFAIRMetrics

Open challenges

- Practices of identifying and locating ‘objects’
  - A data object is assigned with a persistent identifier, which resolves to a landing page that includes metadata and links to access the content.
  - Persistent identifiers for all (data, metadata)
  - Data and metadata in a self-describing format.
- Indicators and priorities may be changed/extended depending on **community practices**, **users** (evaluators) and at which **stage** of the data cycle the assessment is performed.
- Meaningful communication of FAIR assessment results to different stakeholders

icons: https://icons8.com

Thank you for your attention!

Slide acknowledgements:
Ilona von Stein (DANS)
Anusuriya Devaraju (PANGAEA)
Herve L'Hours (UKDA)
Mustapha Mokrane (DANS)

Slide review:
Linas Čepinskas (DANS)

www.fairsfair.eu

@FAIRsFAIR_EU

Resources

- FAIRsFAIR Comments Response on RDA FAIR Data Maturity Model Working Group (2020), https://doi.org/10.5281/zenodo.3827108
- D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures https://doi.org/10.5281/zenodo.3835697
- CoreTrustSeal+FAIR Landscape of Capability Maturity Modeling - A FAIRsFAIR Discussion Paper http://doi.org/10.5281/zenodo.3862588
- FAIRsFAIR Data Objects Assessment Metrics (Version 0.2) http://doi.org/10.5281/zenodo.3775794
- D4.1: Draft recommendations on requirements for FAIR data objects in trustworthy data repositories https://doi.org/10.5281/zenodo.3678716