

# FAIR and TRUST

## The perfect mix

Ingrid Dillo  
Deputy Director DANS

5<sup>th</sup> Joint JST- STM Seminar: The transformation in  
scholarly publishing - Tokyo, 08-11-2022

Data Archiving and Networked Services

**DANS**



# DANS is about keeping data FAIR

Mission:  
enhancing the  
reusability of  
research data and  
thus the quality of  
scientific research

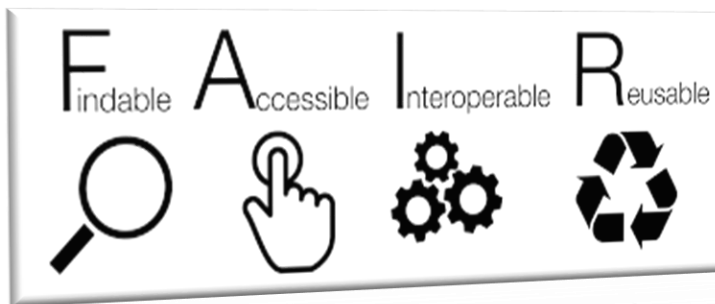
National Centre  
of Expertise  
and Repository  
for Research  
Data

Institute of  
Dutch Academy  
and Research  
Funding  
Organisation  
(KNAW & NWO)  
since 2005

First predecessor  
dates back to  
1964 (Steinmetz  
Foundation),  
Historical Data  
Archive 1989

# Data FAIRPORT(2014)

During a workshop for the life sciences in Leiden in 2014 a minimal set of community-agreed guiding principles were formulated.



# FAIR guiding principles (2016)

The screenshot shows the article page for 'The FAIR Guiding Principles for scientific data management and stewardship' on the Nature Scientific Data website. The page layout includes a top navigation bar with 'nature > scientific data > comment > article' and 'a natureresearch journal'. A secondary bar contains a 'MENU' dropdown and the 'SCIENTIFIC DATA' logo. The article title is prominently displayed, followed by the authors' names. A 'Download PDF' button is visible on the right. Below the title, there are statistics for citations (707) and altmetrics (1322). A 'Sections' sidebar on the right lists various parts of the article, including Abstract, Comment, Additional Information, References, Acknowledgements, Author information, Rights and permissions, and About this article. A note at the bottom indicates an addendum was published on 19 March 2019.

nature > scientific data > comment > article

a natureresearch journal

MENU SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

## The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

Scientific Data 3, Article number: 160018 (2016) | Download Citation

An Addendum to this article was published on 19 March 2019

Search E-alert Submit Login

Download PDF

707 Citations 1322 Altmetric Article metrics >>

Sections References

Abstract

Comment

Additional Information

References

Acknowledgements

Author information

Rights and permissions

About this article

<https://www.nature.com/articles/sdata201618>

# To facilitate re-use data need to be:



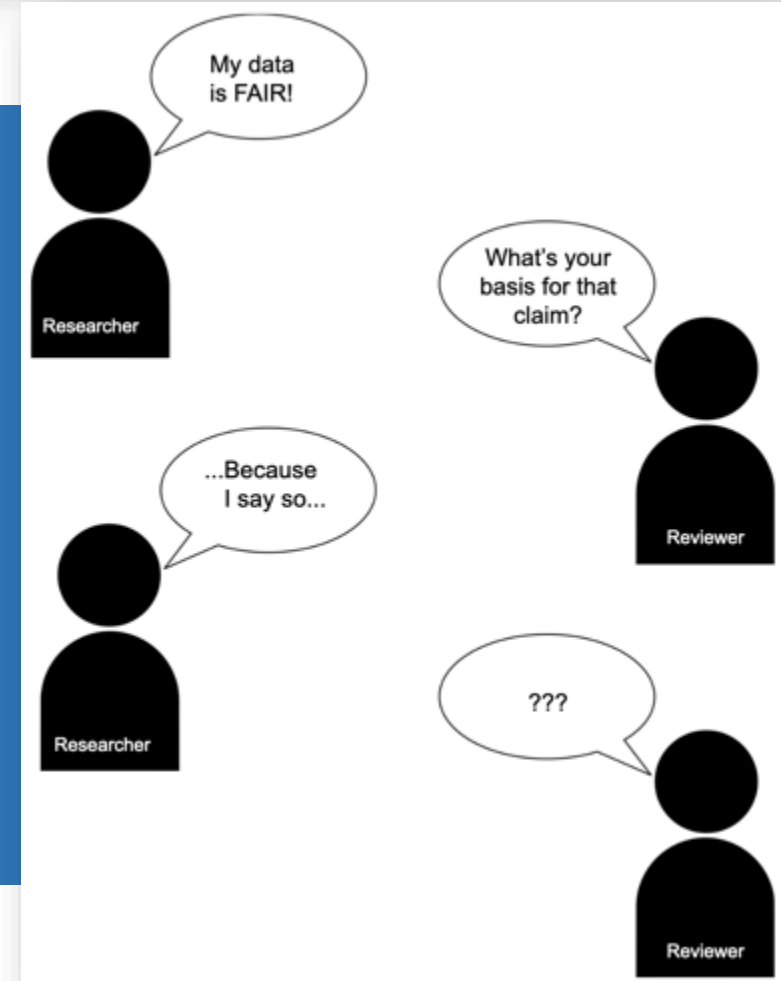
<https://findwise.com/blog/data-that-really-saves-lives-and-possibly-your-organisation/>

# The confusion

continuum

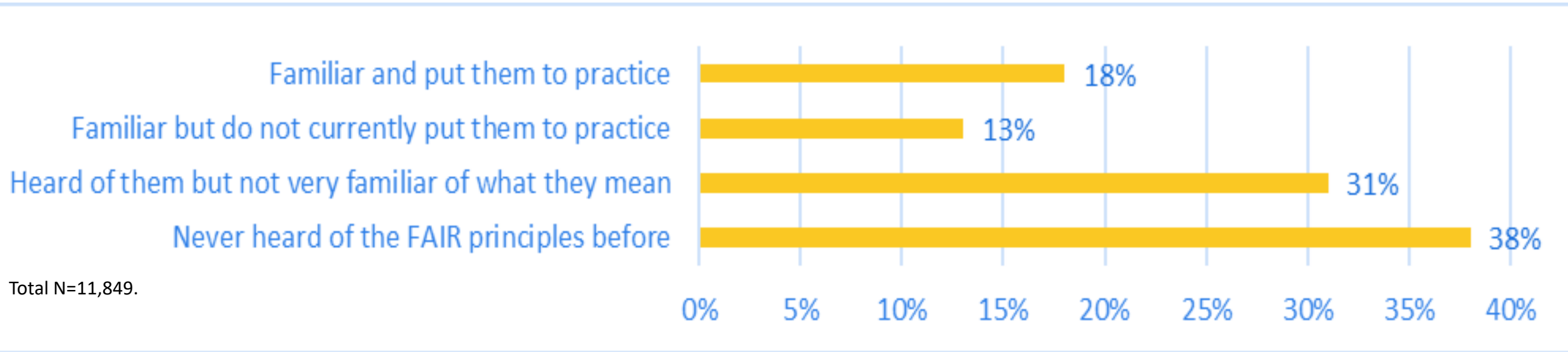
Open data  
is about  
MORE  
THAN  
DISCLOSURE  
it must be  
Fair

- Findable
- Accessible
- Interoperable
- Reusable



# The success: awareness of FAIR

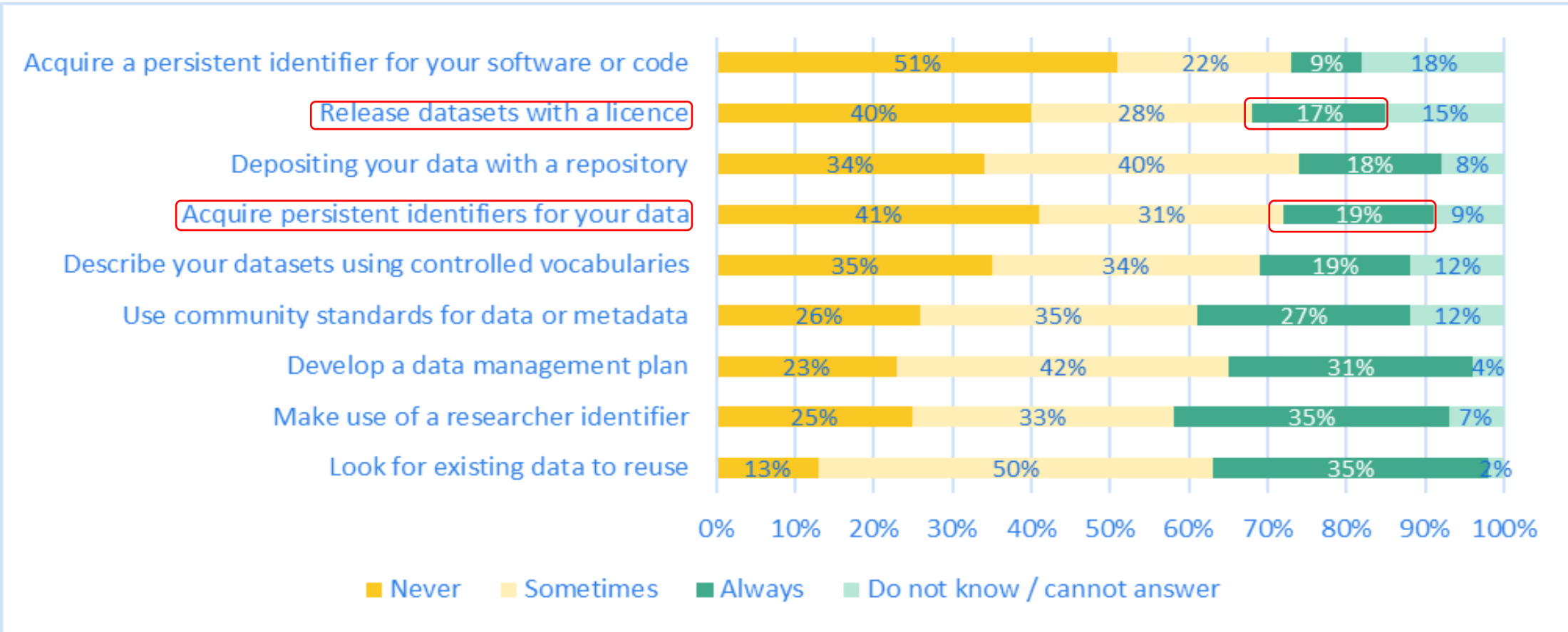
- About two thirds have some level of familiarity with the FAIR principles
- More than a third have never heard of them
- Less than 1 out of 5 puts them into practice



Source: Preliminary findings from the European Research Data Landscape Report (commissioned by the EC)

# The success: FAIR aligned practices

- More than two thirds develop DMPs but other FAIR-aligned practices are less common
- Allocating PIDs to data is the least common practice



# The success

- Well known among policymakers, funders, data service providers
- Less known among researchers



## **Motivators:**

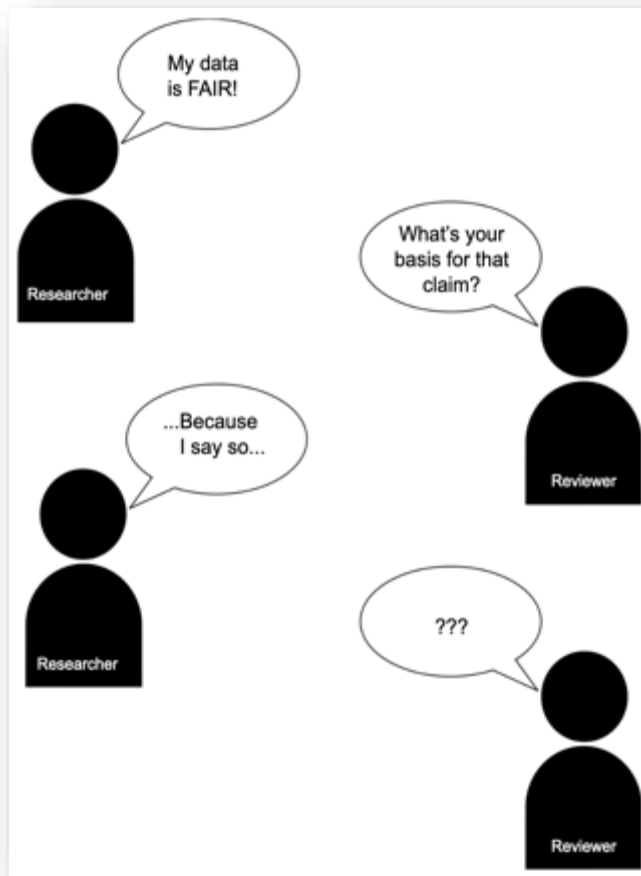
- clear policies
- support for compliance



## **Barriers:**

- time and effort required for RDM and data sharing (academic recognition)
- data protection and legal restrictions

# FAIR metrics and assessment: current status



- many sets of metrics and many tools around
- agreement and convergence is needed
- Formal certification not really on the horizon yet



# FAIR metrics and assessment: challenges



Different assessment tools  $\Rightarrow$  different choices, different implementations of 'weight'  $\Rightarrow$  different scores

How to make sense of assessment scores?

What do principles mean in different research communities?

FAIR principles: <https://force11.org/info/the-fair-data-principles/>  
RDA WG FAIR data maturity model: <https://doi.org/10.15497/RDA00050>  
FAIRsFAIR data object assessment metrics (implemented in [F-UJI tool](#)):  
<https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments>

# Assessment Results:

## Evaluated Resource:

Vertical and horizontal light heterogeneity along with

FAIR level: ?

Resource PID/URL:

DataCite support:

Metric Version:

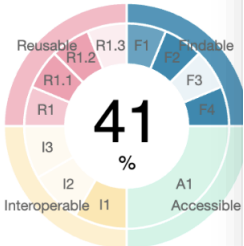
Metric Specification:

Software version:

Download assessment results:

Save and share assessment results:

## Summary:



VERTICAL A  
AND WARM

« Back to list

Overview De

Cite as:

Hiura, Prof. Dr. T  
light heterogenei  
https://doi.org/10

2022-05-24 | Hiura, F  
k5xe

This dataset contains the data on light measures and structural attributes in temperate secondary forests in Japan.

## Interoperable

FsF-I1-01M - Metadata is represented using a formal knowledge representation language.

FsF-I2-01M - Metadata uses semantic resources

FAIR level: 1 of 3  
Score: 0 of 1  
Output: []

initial

Metric tests:	Test:	Test name:	Score:	Maturity:	Result:
	FsF-I2-01M-1	Vocabulary namespace URIs can be identified in metadata	1		✓
	FsF-I2-01M-2	Namespaces of known semantic resources can be identified in metadata			?

## Debug messages:

Level:	Message:
INFO	Number of vocabulary namespaces extracted from all RDF-based metadata -: 2
INFO	Number of vocabulary namespaces extracted from all RDF-based metadata -: 2
INFO	Default vocabulary namespace(s) excluded -: ['http://schema.org']
INFO	Check if the remaining namespace(s) exist(s) in a LOD registry -: ['http://datacite.org/schema']
WARNING	NO vocabulary namespace match is found
WARNING	Vocabulary namespace (s) or URIs specified but no match is found in LOD reference list (examples) -: ['http://datacite.org/schema']

FsF-I2-01M - Metadata includes link between the data and its related entities

# FAIR-Aware tool



Do you work with data? Are you looking to make it future-proof? The **FAIR Principles** can help you. These principles stand for the Findability, Accessibility, Interoperability and Reusability of data(sets) to help others to find, cite and reuse your data more easily.

FAIR-Aware helps you assess your knowledge of the FAIR Principles, and better understand how the value and impact of your data.

The tool is discipline-agnostic, making it relevant to any scientific field. You can use this tool at any stage of your research, from the initial planning of a project to the final publication of your data(sets) in a data repository. It is also good to keep in mind that many FAIR-related decisions can be made so you may want to use FAIR-Aware early on to help you make those decisions. Also, if you are a student, you can use the tool to assess your knowledge of FAIR of your course participants.

The self-assessment consists of 10 questions with additional guidance texts to help you become your data(sets) as FAIR as possible. The assessment will take between 10-30 minutes, after which you will receive additional tips on how you can further improve your FAIR skills.

If you would like to use FAIR-Aware in your own training, you can find instructions on the trainer for trainers page. You can contact the FAIR-Aware development team for any questions or comments via e-mail.

## 6. Are you aware that the metadata describing your data(set) should use controlled vocabularies?

### What does this mean?

There are many different ways you can describe the same information when filling out the metadata for your deposit. To prevent ambiguity and facilitate better findability, interoperability, and machine-readability, you should use a **controlled vocabulary** to enter your metadata.

Controlled vocabularies are lists of terms that are created for specific uses or contexts. They are a type of **semantic artefact** and can take the form of, for example, an ontology, thesaurus, or taxonomy. Each type of vocabulary comes with a different degree of sophistication [e.g. in their level of expressiveness, structure, and inferential power].

### Why is this important?

When using controlled vocabularies, the discovery, linking, understanding, and reuse of research data are improved. Using controlled vocabularies in metadata facilitates enhanced data search because people will not have to guess the exact terms you used to describe your data(set) to find it. It also helps facilitate better interoperability of data from different sources, since it will be clear that data(sets) using the same terms cover the same information.

Data repositories should provide support for the use of controlled vocabularies in metadata by offering relevant functionalities. They will often display which controlled vocabularies they support on their website. When controlled vocabularies are included in the metadata, your data repository of choice may be able to publish the metadata in machine-readable format, thus greatly increasing their machine actionability.

### How to do this?

Controlled vocabularies are often domain-specific. It is recommended to use the vocabulary that is used most often in your field or specific line of research (see Q8). If you are unsure about this, you can contact your research support staff or look up some data(sets) from colleagues in your field.

You can find data repositories supporting your preferred controlled vocabulary in registries such as FAIRsharing or Re3data by filtering on 'metadata standards'. Below is a non-exhaustive list of some registries or look-up services for vocabularies. You can use these resources to search for a vocabulary that covers terms relevant for your research.

- Basel Register of Thesauri, Ontologies & Classifications (BARTOC)
- CESSDA Vocabulary Service
- Linked Open Vocabularies (LOV)
- OBO Foundry
- BioPortal
- NERC Vocabulary Server
- Research Vocabulary Australia
- MMI Ontology Registry and Repository (ORR)
- Industrial Ontologies Foundry (IOF)

### Want to know more?

If your field has no common controlled vocabularies (yet), you can search for one you personally find most suitable. It is recommended to do this in collaboration with your research support staff. Before using a controlled vocabulary, you should establish the following:

- Whether it is available online and is open to other users
- Whether it contains the relevant terms for your line of research
- Whether you know who curates and makes the vocabulary available to other users
- Whether it is an nationally or internationally recognized vocabulary and if it is used extensively

Close

## Questions

Glossary

Are you aware that a data(set) should be assigned a globally unique and resolvable identifier when deposited with a data repository? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that when you deposit a data(set) in a data repository, you should provide discovery metadata in order to make the data findable, understandable and reusable to others? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that the data repository providing access to your data should make the metadata describing your data(set) available to all, including machines as well as humans? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which your data(set) can be reused? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that metadata should remain available over time, even if the data(set) is no longer accessible? <sup>1</sup>

☐ Yes  
☐ No

## Summary

Are you aware that the metadata describing your data(set) should use controlled vocabularies? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that provenance information about the collection and location of data should be included in the metadata? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that metadata describing your data(set) should follow the conventions of a community-endorsed standard? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that your data(set) should be deposited preferably in a data repository that is open and supported by the data repository for long-term preservation? <sup>1</sup>

☐ Yes  
☐ No

Are you aware that keeping your data(set) FAIR over time requires ongoing data curation and digital preservation? <sup>1</sup>

☐ Yes  
☐ No

# FAIRness as a (meta)data snapshot



- Focus on the data and metadata
- Provides a “snapshot” of a digital object in isolation of its context

## FAIRness a FAIRytale?



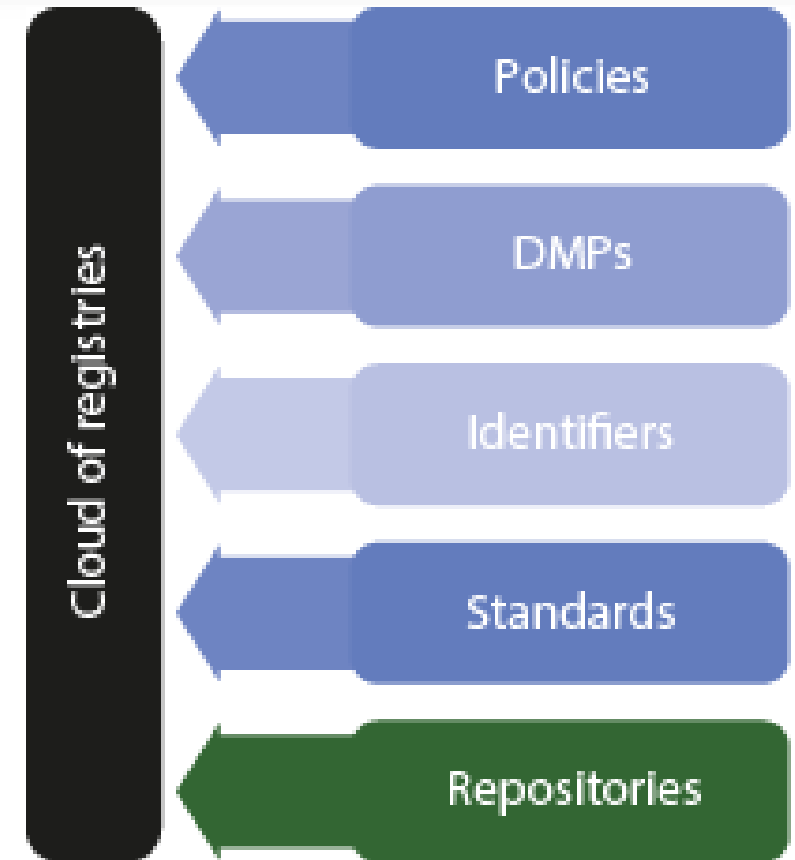
“Research data will not become nor stay FAIR by magic. We need skilled people, transparent processes, interoperable technologies and collaboration to build, operate and maintain research data infrastructures.”

Mari Kleemola, Finnish Social Science Data Archive


<https://tietoarkistoblogi.blogspot.com/2018/11/being-trustworthy-and-fair.html>

# FAIR Data Ecosystem

- **F1.** (meta)data are assigned a globally unique and persistent **identifier**
- **F4.** (meta)data are registered or indexed in a **searchable resource**
- **A1.** (meta)data are retrievable by their identifier using a **standardized communications protocol**
- **A1.2** the protocol allows for an **authentication and authorization** procedure, where necessary
- **A2.** metadata are accessible, even when the **data are no longer available**



*Turning FAIR data into reality, Final report and Action Plan from the European Commission Expert Group on FAIR*  
Data <https://doi.org/10.2777/54599>



“Perhaps the biggest challenge in sharing data is trust: how do you create a system robust enough for scientists to trust that, if they share, their data won’t be lost, garbled, stolen or misused?”

## **The Data Harvest:**

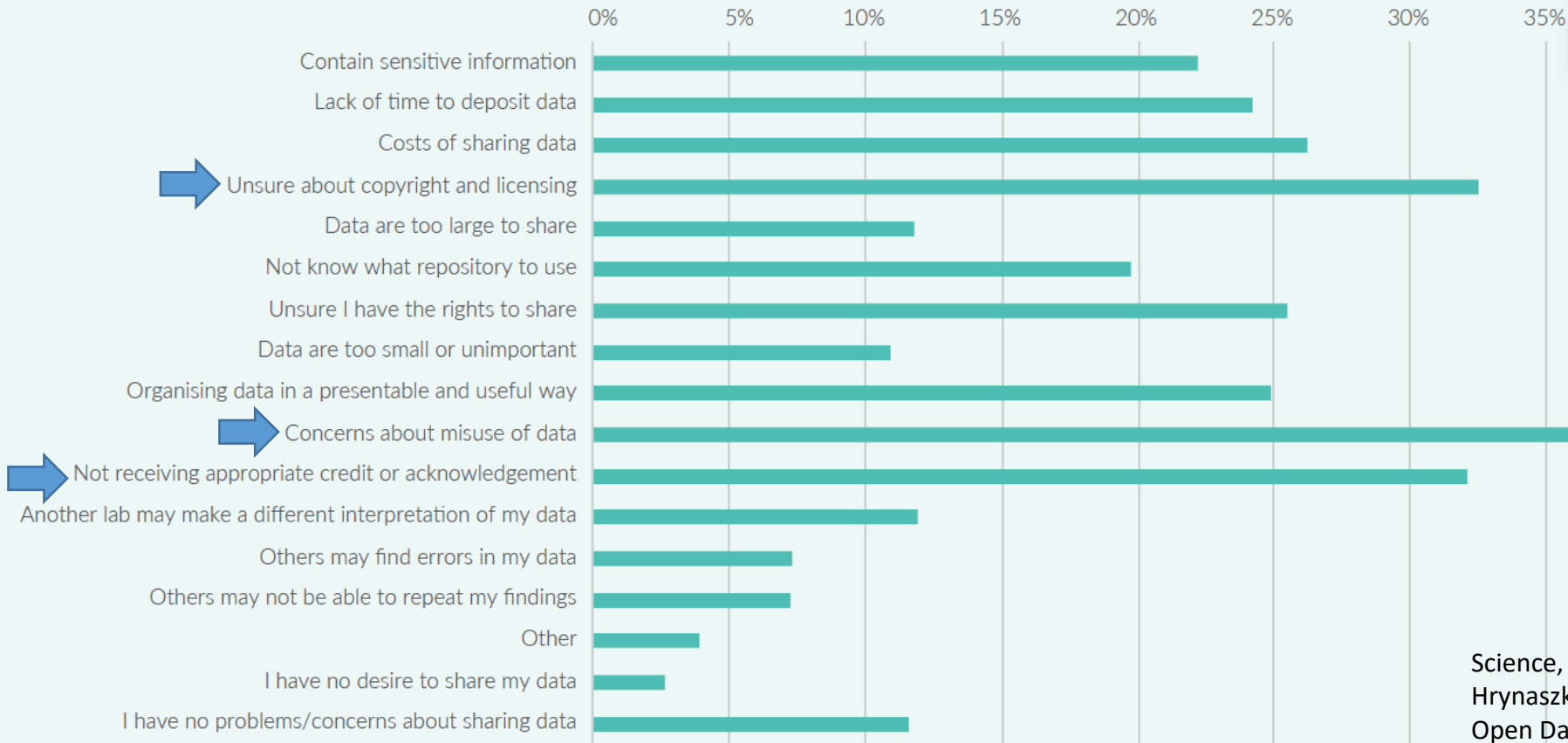
**How sharing research data can yield knowledge, jobs and growth**

**An RDA Europe Report**

**December 2014**

# A matter of trust

Problems/concerns respondents have with sharing datasets



"The biggest barrier to research data sharing and reuse seems to be a matter of trust"

Science, Digital; Fane, Briony; Ayriss, Paul; Hahnel, Mark; Hrynaskiewicz, Iain; Baynes, Grace; et al. (2019): The State of Open Data Report 2019. figshare. Report.  
<https://doi.org/10.6084/m9.figshare.9980783.v2>

# The TRUST principles

- developed in 2019-2020 by the community under the umbrella of the Research Data Alliance
- 19 co-authors representing: 4 continents, diverse stakeholders, multiple scientific domains
- Focus on data repositories
- High level principles to facilitate stakeholder discussion and guide repositories



<https://www.rd-alliance.org/trust-principles-rda-community-effort>

# The TRUST principles



## The TRUST Principles

Principle	Guidance for Repositories
Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Source: Lin et al., 2020. *The TRUST Principles for Digital Repositories*. *Scientific Data*  
<https://doi.org/10.1038/s41597-020-0486-7>

# CoreTrustSeal

- Community driven repository certification standard
- Developed under the umbrella of RDA
- 16 (revised) requirements, reflecting the characteristics of TRUSTworthy Data Repositories (TDRs)

## Minimal (core) standard

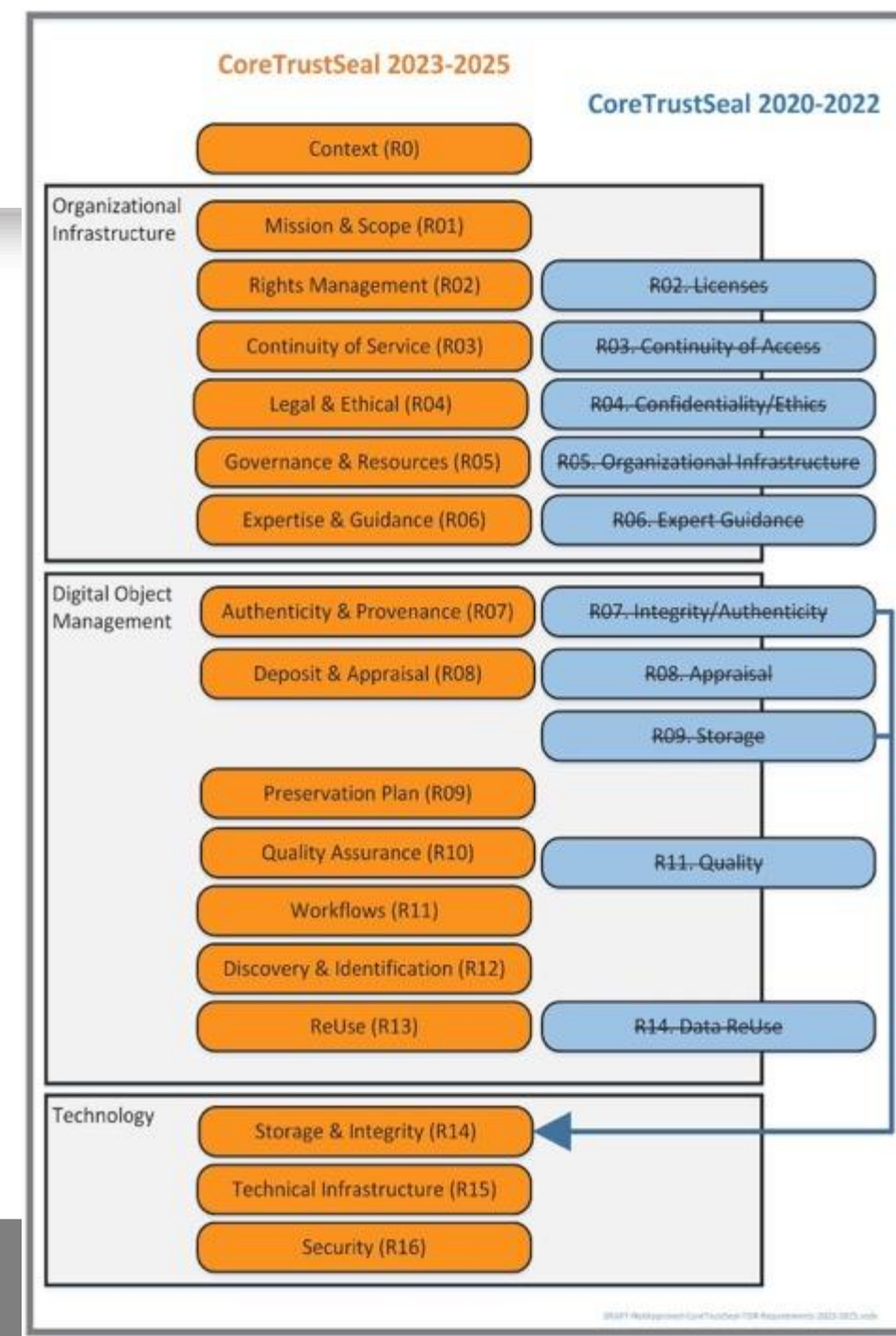
- Self-assessment, peer review, 3 year cycle, transparent processes
- Discipline agnostic, global uptake



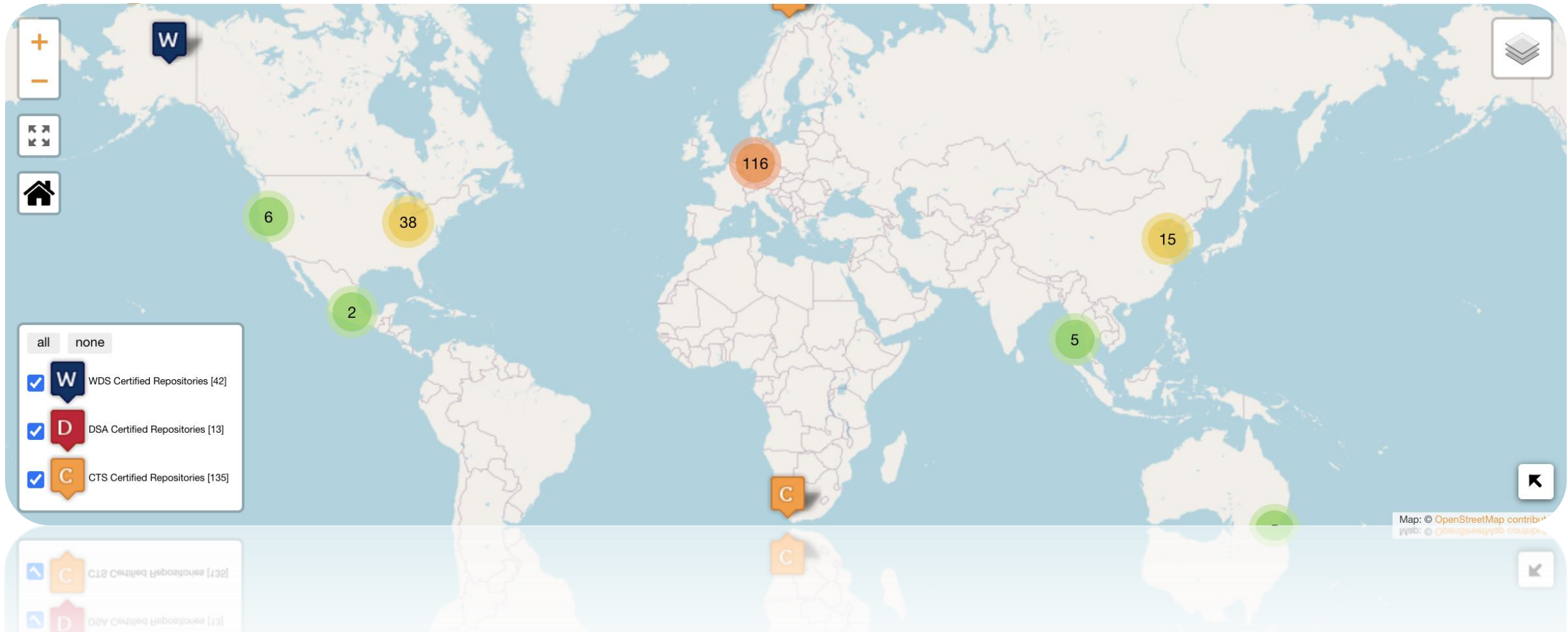
<https://www.coretrustseal.org>

CoreTrustSeal-Requirements-2023-2025\_v01.00

<https://doi.org/10.5281/zenodo.7051095>



# Core Certified Repositories



# CoreTrustSeal: perceived benefits

## External:

- Displays commitment to data and service quality and long-term data curation
- **Heightens stakeholder confidence**
- Increases national and international recognition and reputation
- **Increases your visibility**
- Show data holdings and services are searchable, accessible, and satisfy national and international standards

## Internal:

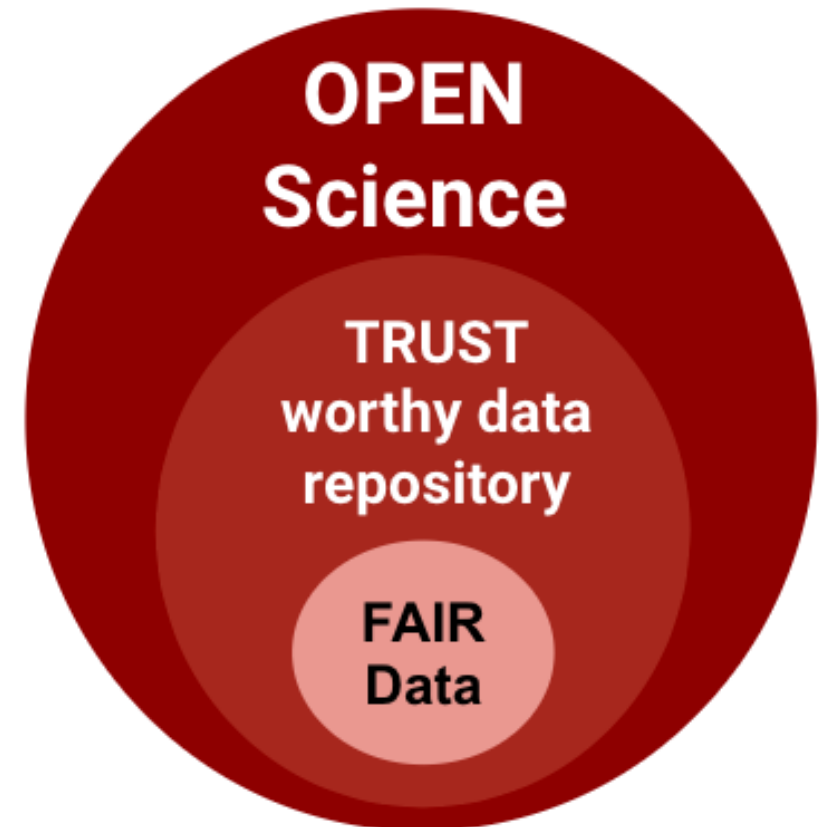
- Benchmark for comparison/ determine strengths and weaknesses
- **Improves professionalism:**
  - Checking, improving and updating policy and workflow documents
  - Re-evaluating and making improvements on our technical solutions and processes for long-term preservation
- Improves awareness and compliance with established standards
- Increases internal communication
- Good team building exercise
- **Ensuring transparency**

# Takeaway message

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;

The TRUST principles and CoreTrustSeal certification help us to create trust in the research data infrastructure we need in order to safeguard the accessibility and assessability of our (FAIR) data for the future.



We foster a FAIR data culture and the uptake of good practices.

[Read our FAIR stories](#)

## Your tailored support for FAIR adoption



Support  
programme  
for Data  
repositories



FAIR  
practices for  
Data  
Stewards



Policy  
enhancement  
support



Handbook &  
good  
practices for  
universities



Resources  
for Research  
Performing  
Organisations

<https://www.fairsfair.eu/>

FAIR-IMPACT open calls

## Adoption & implementation support

FAIR-IMPACT will boost the uptake of FAIR data principles and practices by research performing organisations, data service providers and repositories through a dedicated support programme. Through a series of open calls, interested parties can apply to implement a selection of current tools and methods and receive financial support to enable their participation. We aim to support around 50 organisations over the life of the project ensuring that different domains, geographic areas and stakeholder groups are represented.



<https://fair-impact.eu/>

WP1 PROJECT  
MANAGEMENT,  
SYNCHRONISATION  
AND  
SUSTAINABILITY

WP3  
PERSISTENT  
IDENTIFIERS

WP6 INTEROPERABILITY

WP2 ENGAGEMENT,  
ADOPTION & IMPLEMENTATION

WP4 METADATA  
AND ONTOLOGIES

WP5 METRICS,  
CERTIFICATION  
AND GUIDELINES

WP7  
DISSEMINATION,  
EXPLOITATION  
AND  
COMMUNICATION

**Thank you for listening!**



[ingrid.dillo@dans.knaw.nl](mailto:ingrid.dillo@dans.knaw.nl)  
[www.dans.knaw.nl](http://www.dans.knaw.nl)