



**FAIRsFAIR**  
Fostering Fair Data Practices in Europe

# How good repositories help to make and to keep data FAIR

Marjan Grootveld, Data Archiving and Networked Services (DANS) @MarjanGrootveld

5 March 2020

Workshop «Zertifizierung von Forschungsdatenrepositorien», Leipzig

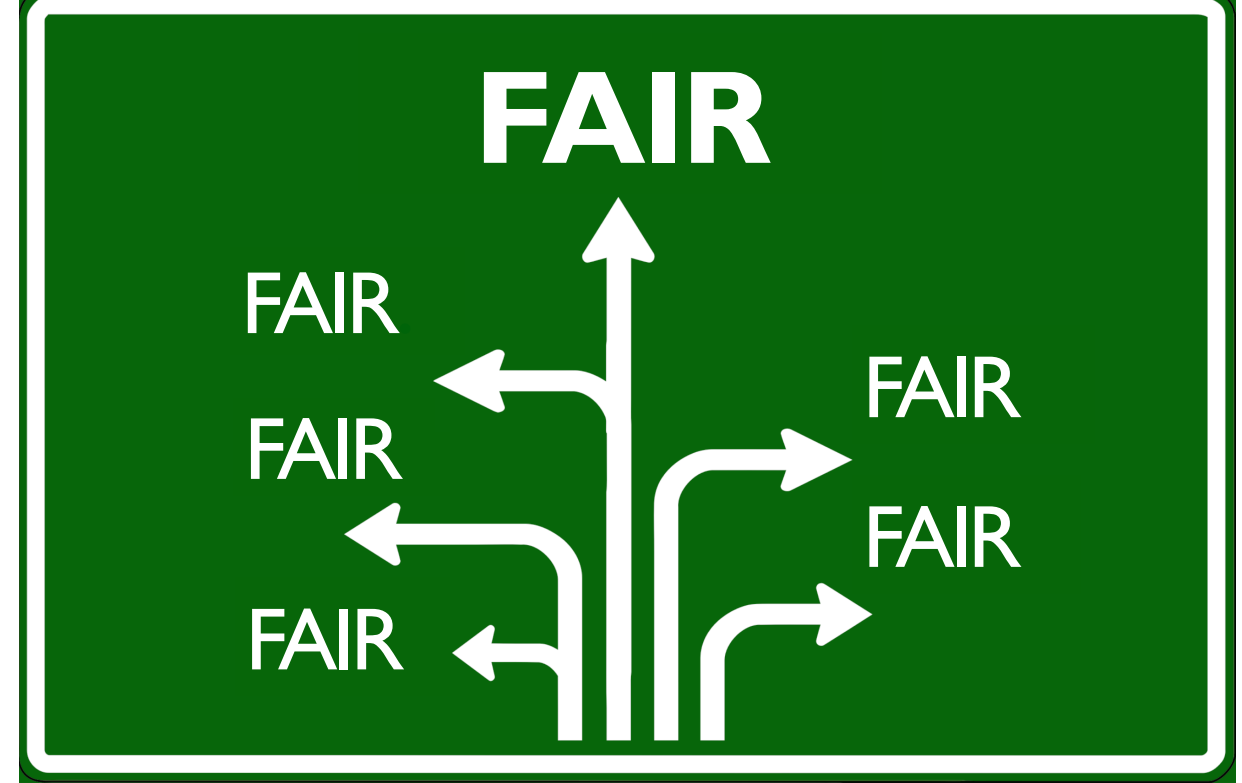


FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

Twitter: @FAIRsFAIR\_EU

## Today's road

- Trust and FAIRness
- Introduction DANS and FAIRsFAIR
- How repositories enable FAIRness
- What's missing in the FAIR principles
- How can we help? And will you win a DANS quartet?



# Core issue in research

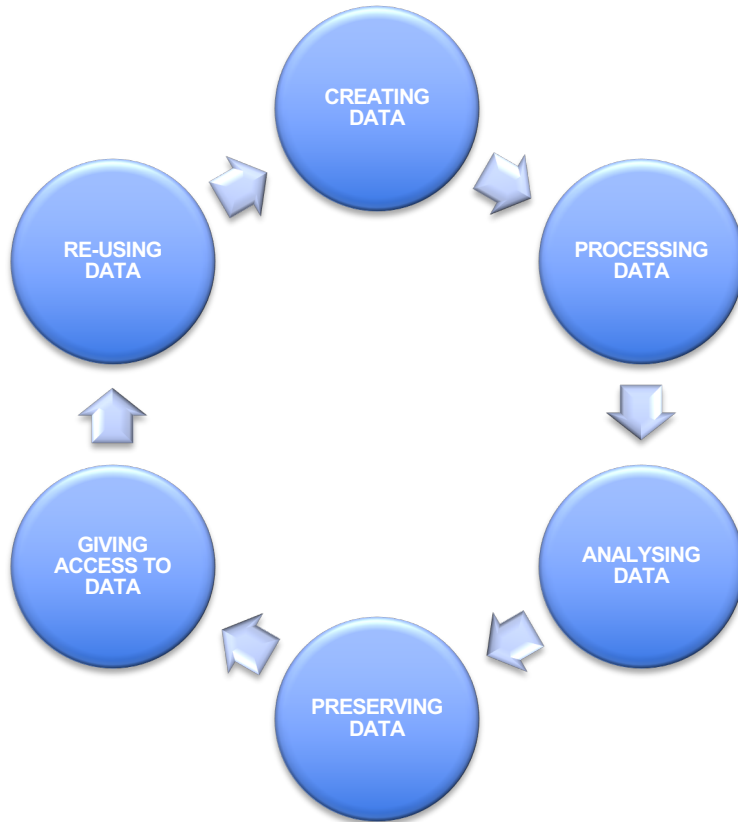


Photo by Yuri Catalano – CC0 - <https://www.pexels.com/photo/city-landscape-sky-people-127420/>

Trust is a central element in the research world:

- Where do these data come from? What has happened with them along the way?
- How can I be sure that “they” interpret and use my data in the right way?

# Simplified research data lifecycle



Managing and documenting data through all stages helps to build trust.

Based on UK Data Archive lifecycle: <https://www.ukdataservice.ac.uk/manage-data/lifecycle>



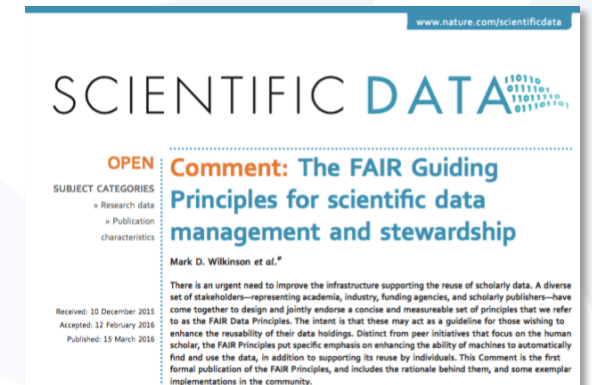




# FAIR data principles

- Findable – Easy to find by **both humans and computer systems** and based on mandatory description of the metadata that allow the discovery of interesting datasets;
- Accessible – Stored for long term such that they can be easily accessed and/or downloaded with **well-defined license and access conditions** (Open Access *when possible*), whether at the level of metadata, or at the level of the actual data content;
- Interoperable – Ready to be combined with other datasets by **humans as well as computer systems**;
- Re-usable – Ready to be used for **future research** and to be processed further **using computational methods**.

<http://www.nature.com/articles/sdata201618>  
[www.force11.org/group/fairgroup/fairprinciples](http://www.force11.org/group/fairgroup/fairprinciples)



# To remember about FAIR

- Principles, not about implementation
- Domain-independent
- Concerns humans and machines
- Most (sub)principles concern data and metadata
- “Accessible” is not the same as “Open”



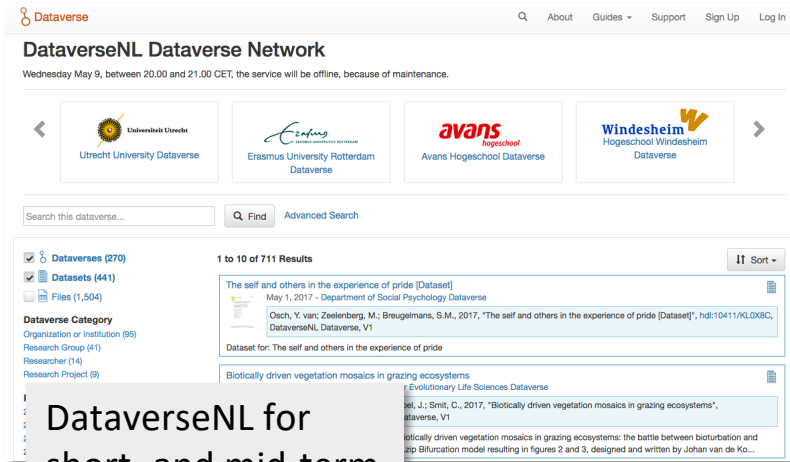
Image: © Marjan Grootveld

# DANS is about keeping data FAIR





# DANS services, see <https://dans.knaw.nl/>



**DataverseNL Dataverse Network**  
Wednesday May 9, between 20.00 and 21.00 CET, the service will be offline, because of maintenance.

Utrecht University Dataverse  
Erasmus University Rotterdam Dataverse  
Avans Hogeschool Dataverse  
Windesheim Hogeschool Windesheim Dataverse

Search this dataverse...

1 to 10 of 711 Results

☒ Datasets (270)  
☒ Datasets (441)  
☐ Files (1,504)

**Dataverse Category**  
Organization or Institution (86)  
Research Group (41)  
Researcher (14)  
Research Project (8)

The self and others in the experience of pride [Dataset]  
May 1, 2017 - Department of Social Psychology Dataverse  
Osch, Y. van; Zaelenberg, M.; Breugelmans, S.M., 2017, "The self and others in the experience of pride [Dataset]", hdi:10411/KL0XSC, DataverseNL Dataverse, V1  
Dataset for: The self and others in the experience of pride  
Biotically driven vegetation mosaics in grazing ecosystems  
Evolutionary Life Sciences Dataverse  
Smit, J.; Smit, C., 2017, "Biotically driven vegetation mosaics in grazing ecosystems", dataverse, V1  
Biotically driven vegetation mosaics in grazing ecosystems: the battle between bioturbation and  
zip Bifurcation model resulting in figures 2 and 3, designed and written by Johan van de Ko...

DataverseNL for  
short- and mid-term  
data storage



**DANS** **NARCIS**

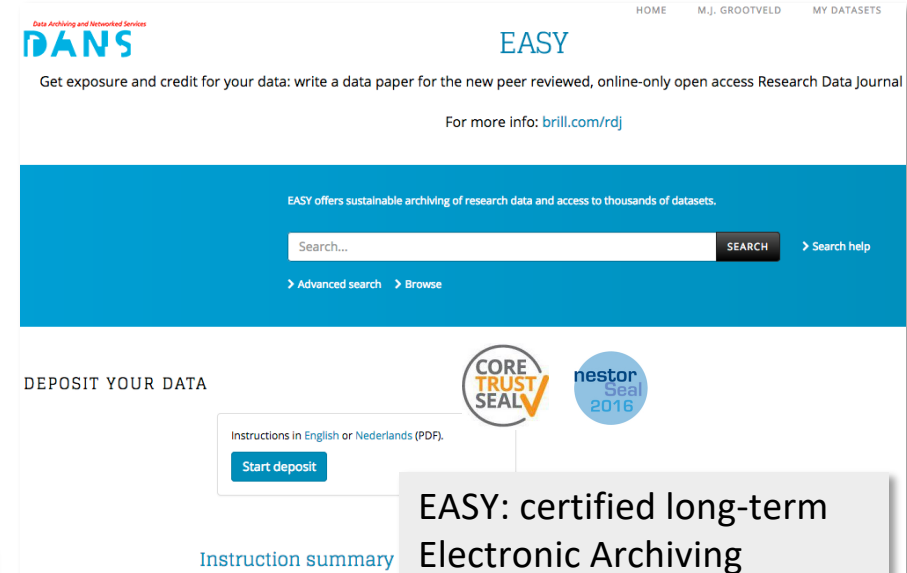
The gateway to scholarly information in the Netherlands

Submit Content to NARCIS

Search...

1,941,501 PUBLICATIONS  
228,629 DATA SETS  
69,927 RESEARCH  
60,827 PEOPLE  
2,971 ORGANISATIONS

NARCIS: Gateway  
to scholarly  
information in  
the Netherlands



**DANS** **EASY**

Get exposure and credit for your data: write a data paper for the new peer reviewed, online-only open access Research Data Journal

For more info: [brill.com/rdj](http://brill.com/rdj)

EASY offers sustainable archiving of research data and access to thousands of datasets.

Search...   [Search help](#)

Advanced search Browse

DEPOSIT YOUR DATA

Instructions in English or Nederlands (PDF).

Instruction summary

CORE TRUST SEAL  
nestor Seal 2016

EASY: certified long-term  
Electronic Archiving  
System for self-deposit

Benefit from our knowledge on research data management  
by our **training sessions, consultancy** and information material.



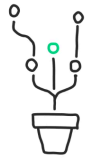
**TRAINING**  
DANS supports researchers (indirectly) in data management by providing training sessions. »

**CONSULTANCY**  
DANS assists in developing data management policy and obtaining certification. »

**INFORMATION MATERIAL**  
Watch the video 'Why share data' or download other information material. »

# FAIR software route

**#1**  
**USE A PUBLICLY**  
**ACCESSIBLE**  
**REPOSITORY**  
**WITH VERSION**  
**CONTROL**



**#3**  
**REGISTER YOUR**  
**CODE IN A**  
**COMMUNITY**  
**REGISTRY**



**#5**  
**USE A SOFTWARE**  
**QUALITY**  
**CHECKLIST**



**#2**  
**ADD A LICENSE**



**#4**  
**ENABLE CITATION**  
**OF THE**  
**SOFTWARE**



Collaboration between DANS and The Netherlands eScience Center – everyone welcome to endorse the ambition and the steps.

<https://fair-software.nl/>

# FAIRsFAIR project

The **FAIRsFAIR-project** aims to supply practical solutions for the use of FAIR data principles

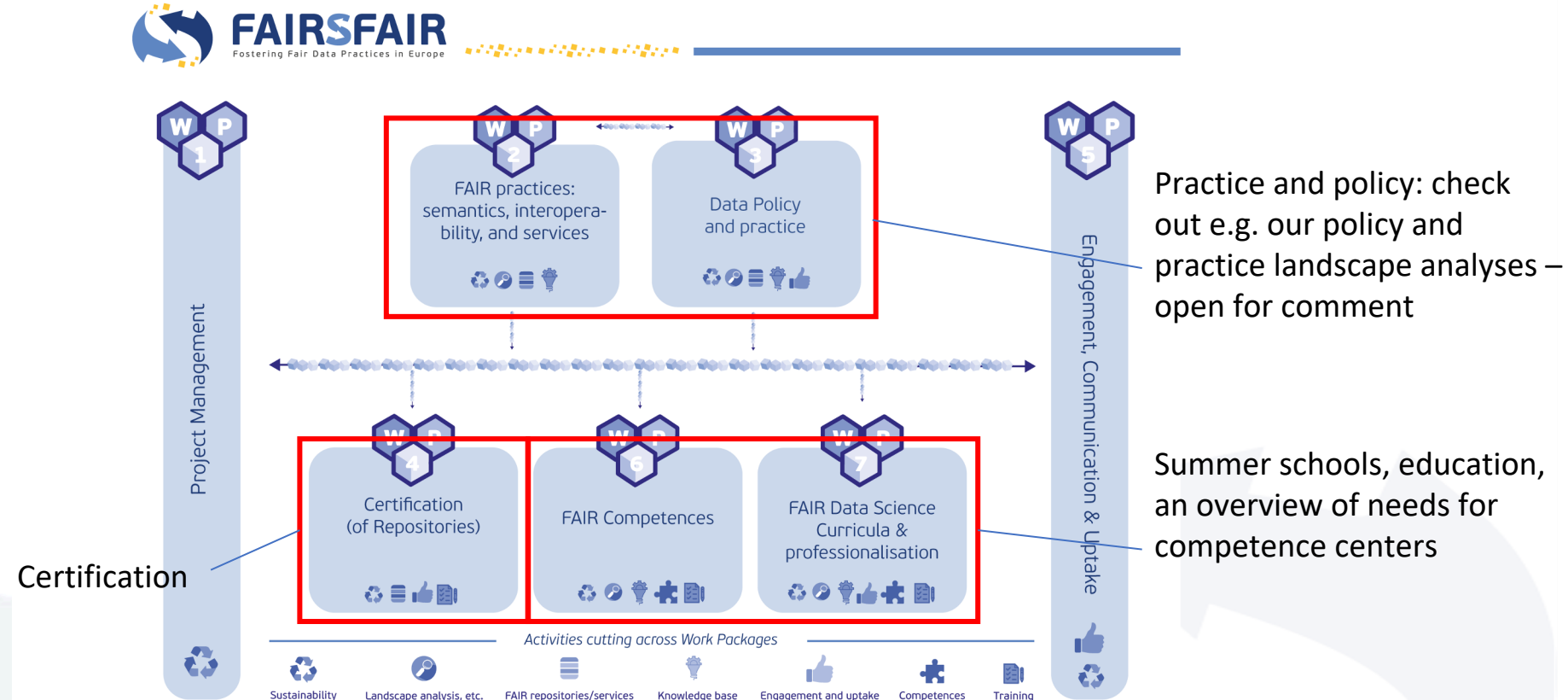
March 2019 – Feb 2022

22 partners

6 core partners DANS (coordinator), CSC, DCC, TrustIT, STFC, EUA

[www.fairsfair.eu](http://www.fairsfair.eu)

@FAIRsFAIR\_EU



# FAIRsFAIR and assessment of FAIR-enabling repositories

- Support the **FAIR alignment of certification schemes** for data repositories, building on existing frameworks such as CoreTrustSeal
- Call for repositories involvement – in depth FAIR-aligned certification support - to expand the **European network of trustworthy repositories enabling FAIR data**
- Provide an improved **registry for finding** and selecting relevant trustworthy **repositories**





# FAIRsFAIR recommends researchers: Deposit your data in a **trustworthy** repository

- ✓ Certified repositories are assessed against a set of guidelines to evaluate their **trustworthiness**
- ✓ A **certified** repository typically provides several services that ensure the **FAIRness of your dataset**.
- ✓ Certification frameworks exist that can be used to assess the quality of a repository
- ✓ CoreTrustSeal is in common use for this



# How do repositories **make** data FAIR?

For example:

- by providing **persistent unique identifiers**
  - long-term findability, sustainable citations, appropriate academic credit
- by supporting findability through a **public catalogue**
  - effective data discovery is key to data sharing
- by supporting depositors to add a **usage licence** for the data
  - clear terms and conditions that meet legal requirements
- by implementing and promoting **metadata standards**
  - interoperability



Digitalbevaring.dk



Digitalbevaring.dk

Illustrations: Jørgen Stamp:  
Digitalbevaring (CC-BY  
from <http://digitalbevaring.dk/digital-bevaring/>)

See also "Top 10 FAIR Data & Software Things". Zenodo. <http://doi.org/10.5281/zenodo.3409968>

# How do repositories **keep** data FAIR?

They provide the **long-term stewardship** of FAIR digital objects, including **curation activities**, to ensure that the **data remains FAIR over time**

- Support for data producers (e.g. on file formats)
- Support for data users (e.g. on citation)

See also Mokrane & Recker, 2019. CoreTrustSeal-certified repositories. Enabling Findable, Accessible, Interoperable and Reusable (FAIR) data.  
[https://ipres2019.org/static/pdf/iPres2019\\_paper\\_74.pdf](https://ipres2019.org/static/pdf/iPres2019_paper_74.pdf)



Image: <https://datasupport.researchdata.nl>

## Tools to help assess the FAIRness of data

- The RDA FAIR Data Maturity Model WG provides a nice [overview of FAIR assessment tools \(10+\)](#)
  - develops a set of core assessment criteria
  - no definition on how to evaluate the core criteria
- FAIR assessment tools:
  - serve a variety of motivations, scenarios and approaches
  - serve various stages of the research data life cycle
  - serve different stakeholders





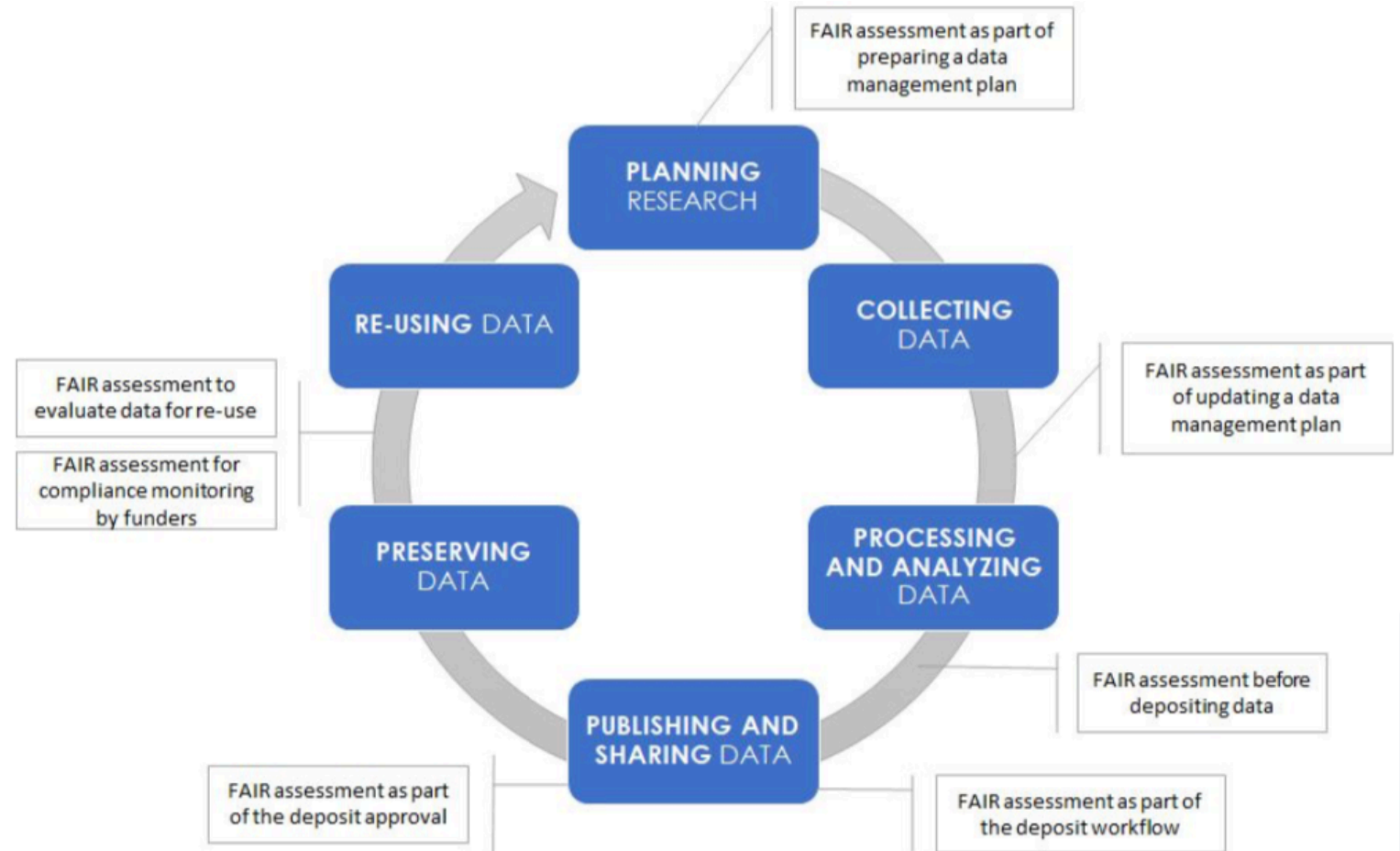
# CoreTrustSeal-FAIR mapping



See Mokrane & Recker, 2019. CoreTrustSeal-certified repositories.  
 Enabling Findable, Accessible, Interoperable and Reusable (FAIR) data.  
[https://ipres2019.org/static/pdf/iPres2019\\_paper\\_74.pdf](https://ipres2019.org/static/pdf/iPres2019_paper_74.pdf) and <https://osf.io/9da2x/>

# FAIRsFAIR assessment of data objects in TDRs

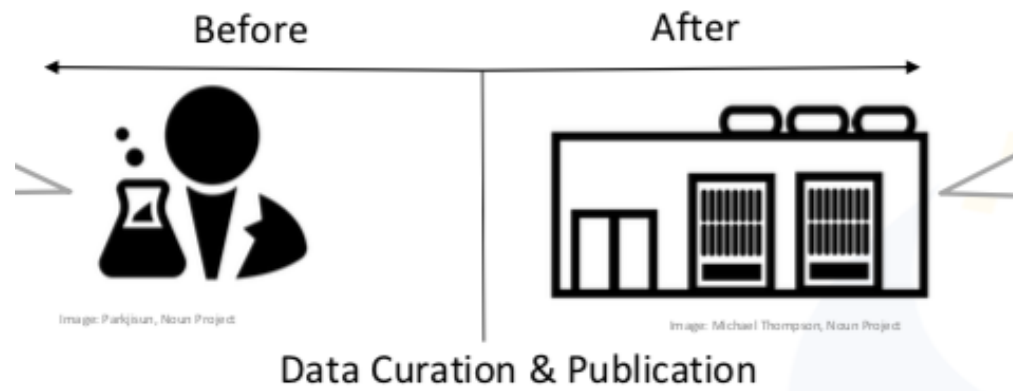
- **Data objects** are individual research datasets that are measured, collected, or created for scientific analysis.
- **Trustworthy digital repository** refers to a certified data repository, e.g., CoreTrustSeal.



A. Devaraju and P. Herterich: D4.1 Draft Recommendations on Requirements for Fair Datasets in Certified Repositories. Zenodo: <https://doi.org/10.5281/zenodo.3678716>

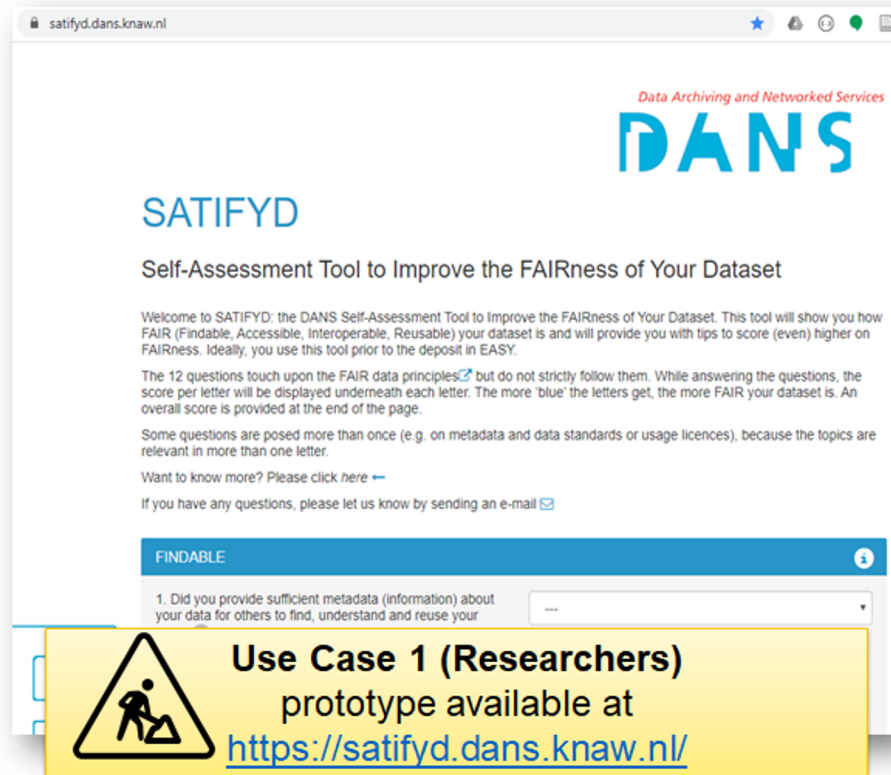
# FAIRsFAIR approach to assessment of data FAIRness

- Use-case based approach:
  - manual self-assessment – **researchers** – prior to deposit
  - automatic assessment – **data repository** – published data sets

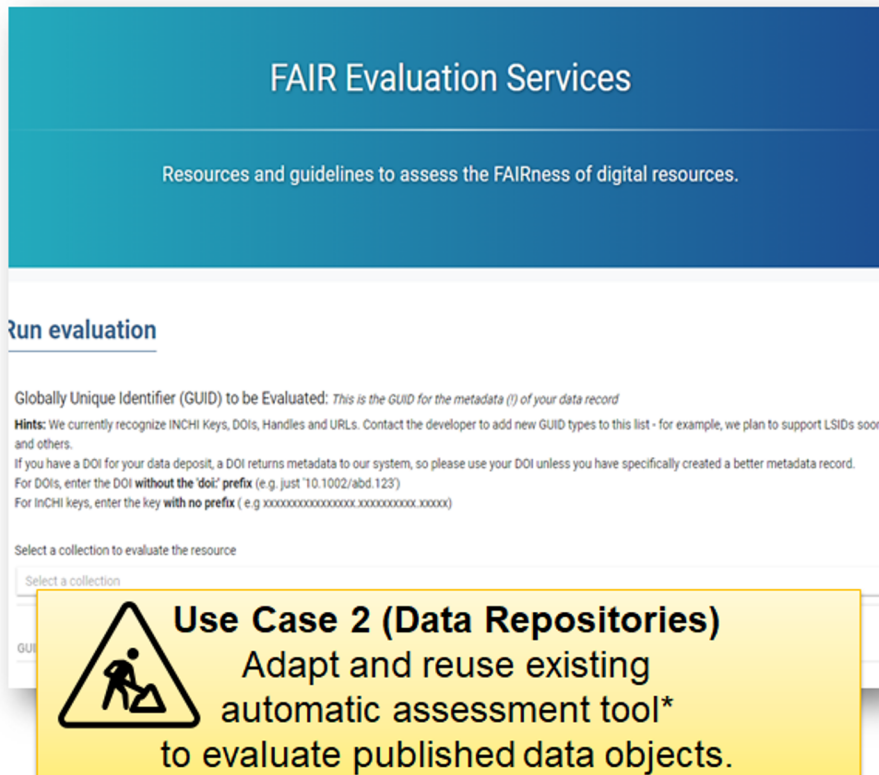


# Two assessment pilots

\* Wilkinson, Mark D., Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, et al. 2019. 'Evaluating FAIR Maturity through a Scalable, Automated, Community-Governed Framework'. Scientific Data 6 (1): 1–12. <https://doi.org/10.1038/s41597-019-0184-5>



The screenshot shows the SATIFYD website interface. At the top, it says "Data Archiving and Networked Services" and "DANS". Below that is the "SATIFYD" logo and the title "Self-Assessment Tool to Improve the FAIRness of Your Dataset". The main text welcomes users and explains the tool's purpose. A section titled "FINDABLE" contains a question: "1. Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your". A yellow callout box at the bottom left contains a warning icon and text: "Use Case 1 (Researchers) prototype available at <https://satifyd.dans.knaw.nl/>".



The screenshot shows the FAIR Evaluation Services website. The header says "FAIR Evaluation Services" and "Resources and guidelines to assess the FAIRness of digital resources." Below that is a section titled "Run evaluation". The text explains how to use the tool, including a "Hints" section and a "Select a collection to evaluate the resource" dropdown menu. A yellow callout box at the bottom right contains a warning icon and text: "Use Case 2 (Data Repositories) Adapt and reuse existing automatic assessment tool\* to evaluate published data objects."



## «What does FAIR miss?»

Previous slides mapped CoreTrustSeal requirements onto FAIR principles. However, there are “extra” CTS requirements, e.g.

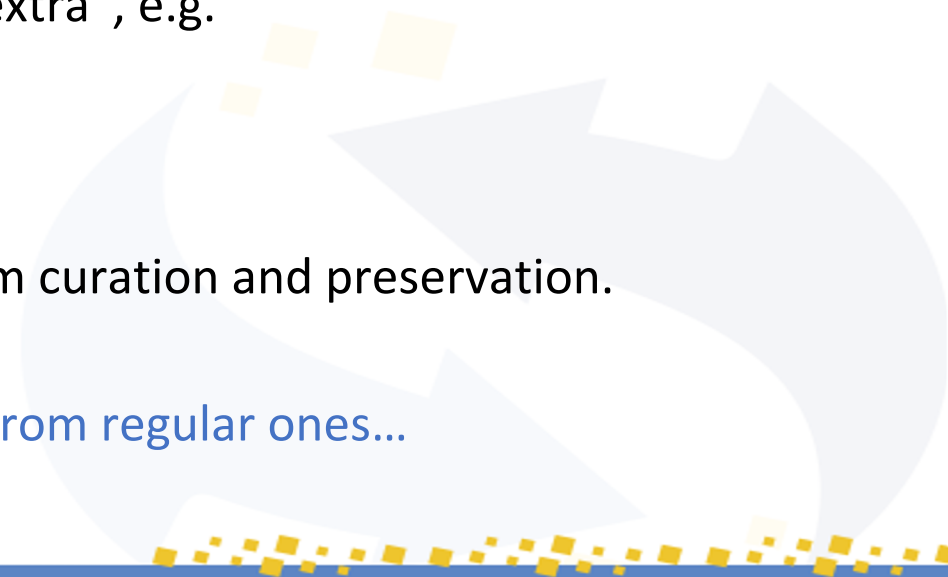
- R0. Designated community
- R4. Confidentiality/Ethics
- R5. Organizational infrastructure
- R6. Expert guidance
- R12. Workflows

And if you map the nestor criteria to FAIR, even more criteria are “extra”, e.g.

- C13. Significant properties
- C14-16. Integrity
- C25. Transformation from AIP to DIP

In short: domain/community, repository organisation, and long-term curation and preservation.

Advocating this is essential to distinguish **trustworthy** repositories from regular ones...



«What does FAIR miss?»

***FAIR for the Future***

***Be interoperable with next generation***

***Trust over Time***

# How can we help you?

What should FAIRsFAIR do to support repositories in supporting FAIR research data:

- A. Give more information about the FAIR principles
- B. Build an assessment tool for the FAIRness of data in your repository
- C. Build a benchmarking tool for the FAIRness of your repository
- D. Anything else?



Slide acknowledgements:

Ilona von Stein (DANS)

Mustapha Mokrane (DANS)

Herve L'Hours (UKDA)

Anusuriya Devaraju (PANGAEA)

[marjan.grootveld@dans.knaw.nl](mailto:marjan.grootveld@dans.knaw.nl)

Twitter: @FAIRsFAIR\_EU

[www.fairsfair.eu/fairsfair-newsletters](http://www.fairsfair.eu/fairsfair-newsletters)