



FAIRsFAIR
Fostering Fair Data Practices in Europe

Fostering FAIR Data Practices in Europe: FAIRsFAIR and the role of repositories

Ingrid Dillo

Deputy director DANS/Project Coordinator FAIRsFAIR
STM Multi-stakeholder webinar on Research Data
19-05-2020





1964: The Steinmetz archive is founded on 27 November. This initiative can be considered the starting point of data archiving in the humanities and social sciences in the Netherlands.

1994: The Scientific Statistical Agency (WSA) is established with the aim of obtaining large data files from organisations such as Statistics Netherlands (CBS).

2005: On 1 June, Data Archiving and Networked Services (DANS) is launched. The Steinmetz, NHDA, WSA and EDNA archives are merged into DANS, which starts to develop various new facilities.

2011: The National Academic Research and Collaborations Information System (NARCIS) is included as a DANS facility. The NARCIS database was developed to enhance the visibility and findability of Dutch research.

2020: DANS celebrates its 15th anniversary!



1987: The Dutch Historical Data Archive (NHDA) is initiated by the History and Informatics working group of Leiden University.

2004: Launch of the E-depot of Dutch Archaeology (EDNA), an archive for electronic archaeological excavation data.

2007: This year sees the launch of the Electronic Archiving System (EASY), the number one online service for depositing and downloading scientific research data.

2014: Management of DataverseNL, a network of data repositories for storing, sharing and publishing research data, is taken over by DANS.

Some figures

EASY: 120,000+ datasets

NARCIS: 265,000+ datasets, 2,000,000+ publications, and more

DataverseNL: 1,300+ datasets and 37,500+ downloads

Training & Consultancy: 1,000+ trainees and 50+ consultancy projects

STAY UP TO DATE
DANS communicates through these channels



Twitter
@DANSKNAW
1,250+ followers



DataLink
digital newsletter
5,000+ subscribers



E-data & Research
magazine
9,800+ subscribers



Website:
dans.knaw.nl
83,000+ visitors

- European Open Science Cloud (EOSC)
- European FAIRsFAIR project
- Trustworthy Digital Repositories (TDRs)



European Open Science Cloud

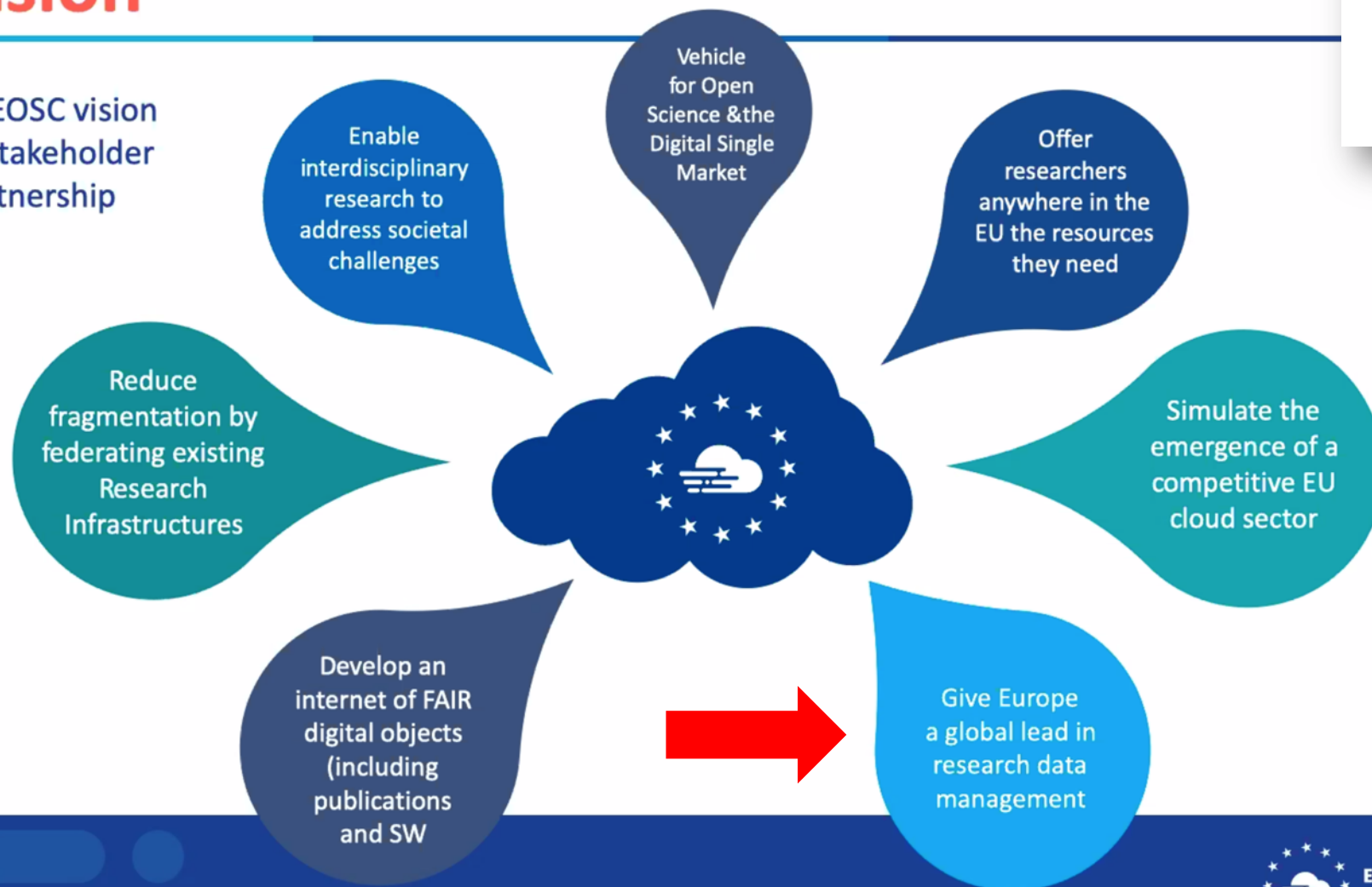
The idea of a EOSC took shape in 2015, as vision of the EC of a large infrastructure to support and develop **open science** and open innovation in Europe and beyond.

The EOSC will be Europe's virtual environment for all researchers to store, manage, analyse and re-use **data** for research, innovation and educational purposes




The Vision

Enabling the EOSC vision
with a multi-stakeholder
European partnership





A photograph of Ursula von der Leyen, a woman with short blonde hair, wearing a red jacket over a white collared shirt. She is speaking and gesturing with her hands. The background is blurred, showing an outdoor setting with other people.

Ursula von der Leyen

World Economic Forum - Davos
22 January 2020

“We are creating the European Open Science Cloud now. It is a trusted space for researchers to store their data and to access data from researchers from other disciplines. We will create a pool of information leading to a web of research insight.”

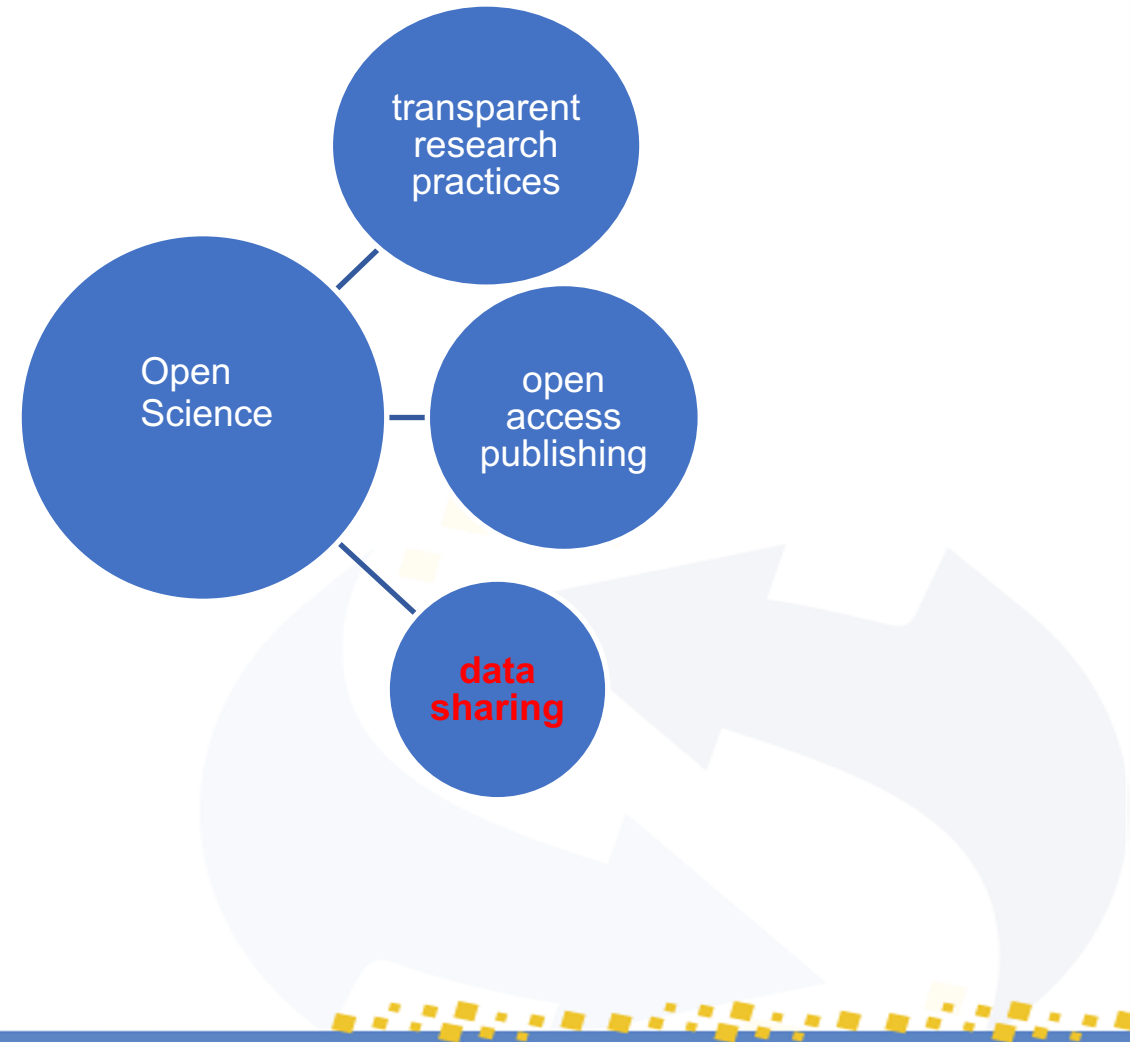
Components of Open Science

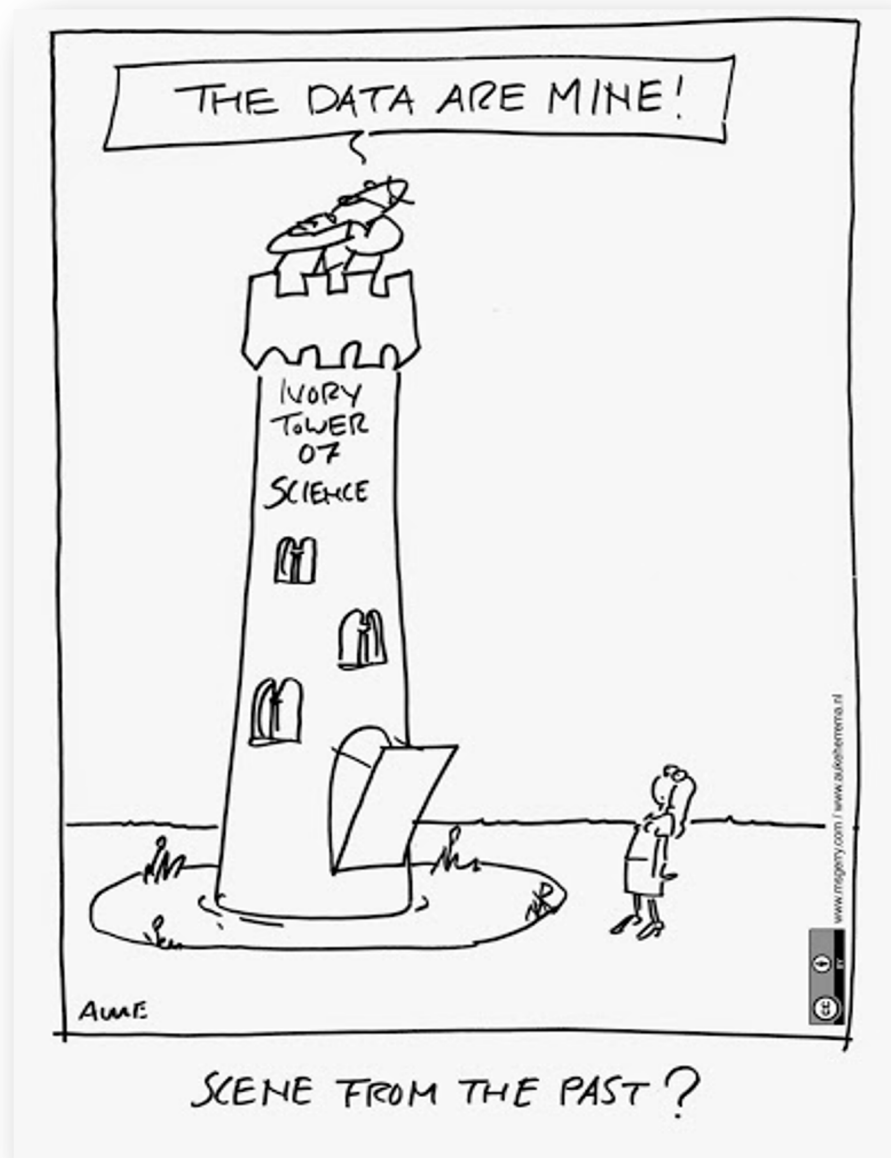
Open science



Open science is an umbrella term for transparent science with ease of access to all products from beginning to end

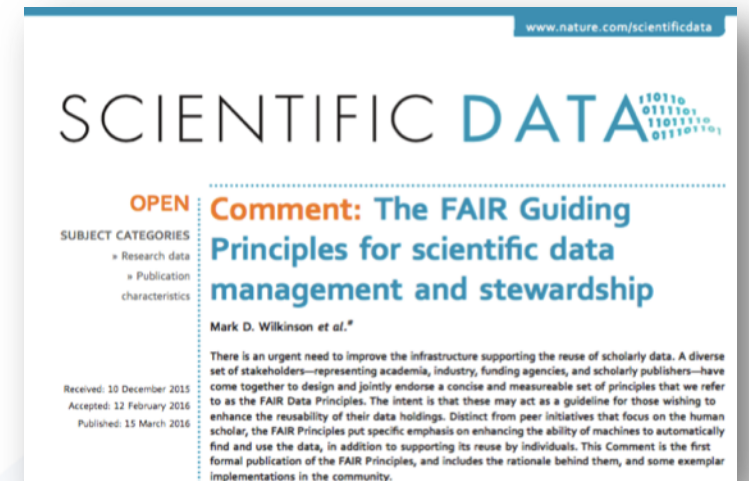
Image credit: Gema Bueno de la Fuente by CC-BY





FAIR 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 licence 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.dtls.nl/fair-data/>
www.force11.org/group/fairgroup/fairprinciples
<http://www.nature.com/articles/sdata201618>

The concept of FAIR data: what does it really mean?

continuum

Open
data
is about
MORE
THAN
DISCLOSURE
it must be
“Fair”

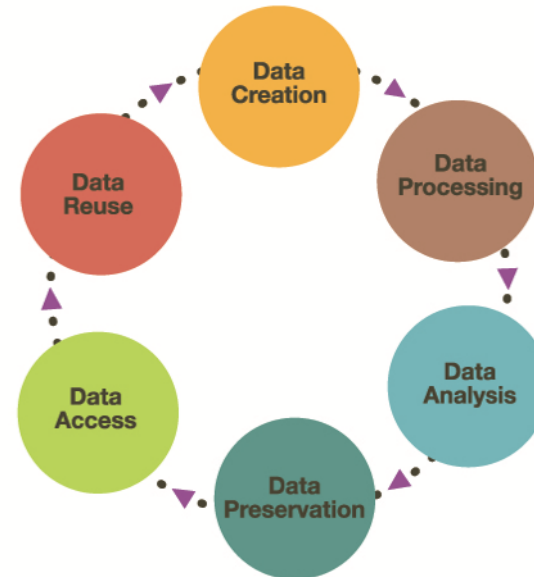
- Findable
- Accessible
- Interoperable
- Reusable



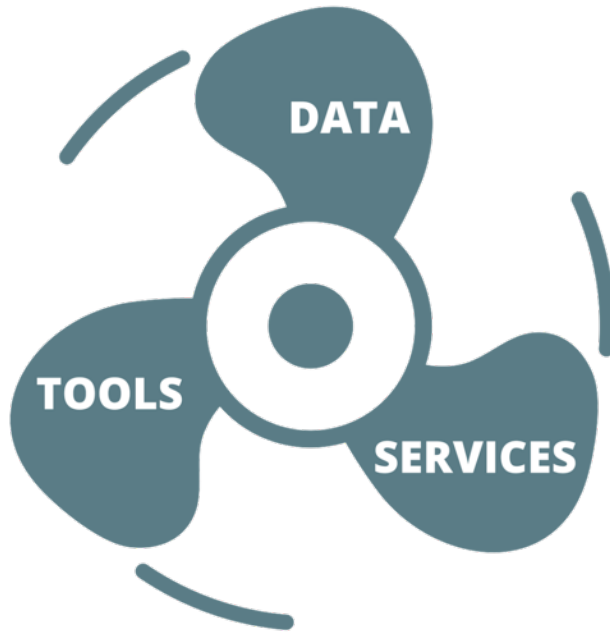
Responsible Research Data Management



PUBLICATIONS AND DATA



FAIR and Europe



https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf



Key Points: To make FAIR a reality ...

- Report takes a holistic approach, not a data centric approach
- Need to address the enabling practices and technologies – not just focus on the data and its attributes
- Need to consider all digital outputs (data, code, metadata etc)
- Objective is to make data and other digital research outputs FAIR for humans and machines.
- Needs: concept of FAIR digital objects, FAIR ecosystem, interoperability frameworks for disciplines and across disciplines, FAIR services including trusted digital repositories, skills, metrics and sustainable funding.

FAIRsFAIR in a nutshell



- supplying practical solutions for the use of the FAIR data principles throughout the research data life cycle;
- fostering FAIR data culture and the uptake of good practices in making data FAIR;
- key role in the development of standards for FAIR certification of repositories and the data within them.

**FAIRSFAR**
Fostering Fair Data Practices in Europe

About

FAIR Support


FAIR Landscape

Our Champions

Events

Project Outputs

Outreach



Health, climate, social science: the “once in a lifetime” moment for data repositories to address the COVID-19 emergency

[Read the article](#)

FAIRsFAIR - Fostering Fair Data Practices in Europe - aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. Emphasis is on fostering FAIR data culture and the uptake of good practices in making data FAIR. FAIRsFAIR will play a key role in the development of global standards for FAIR certification of repositories and the data within them contributing to those policies and practices that will turn the EOSC programme into a functioning infrastructure.

In the end, FAIRsFAIR will provide a platform for using and implementing the FAIR principles in the day to day work of European research data providers and repositories. FAIRsFAIR will also deliver essential FAIR dimensions of the Rules of Participation (RoP) and regulatory compliance for participation in the EOSC. The EOSC governance structure will use these FAIR aligned RoPs to establish whether components of the infrastructure function in a FAIR manner.

www.fairsfair.eu



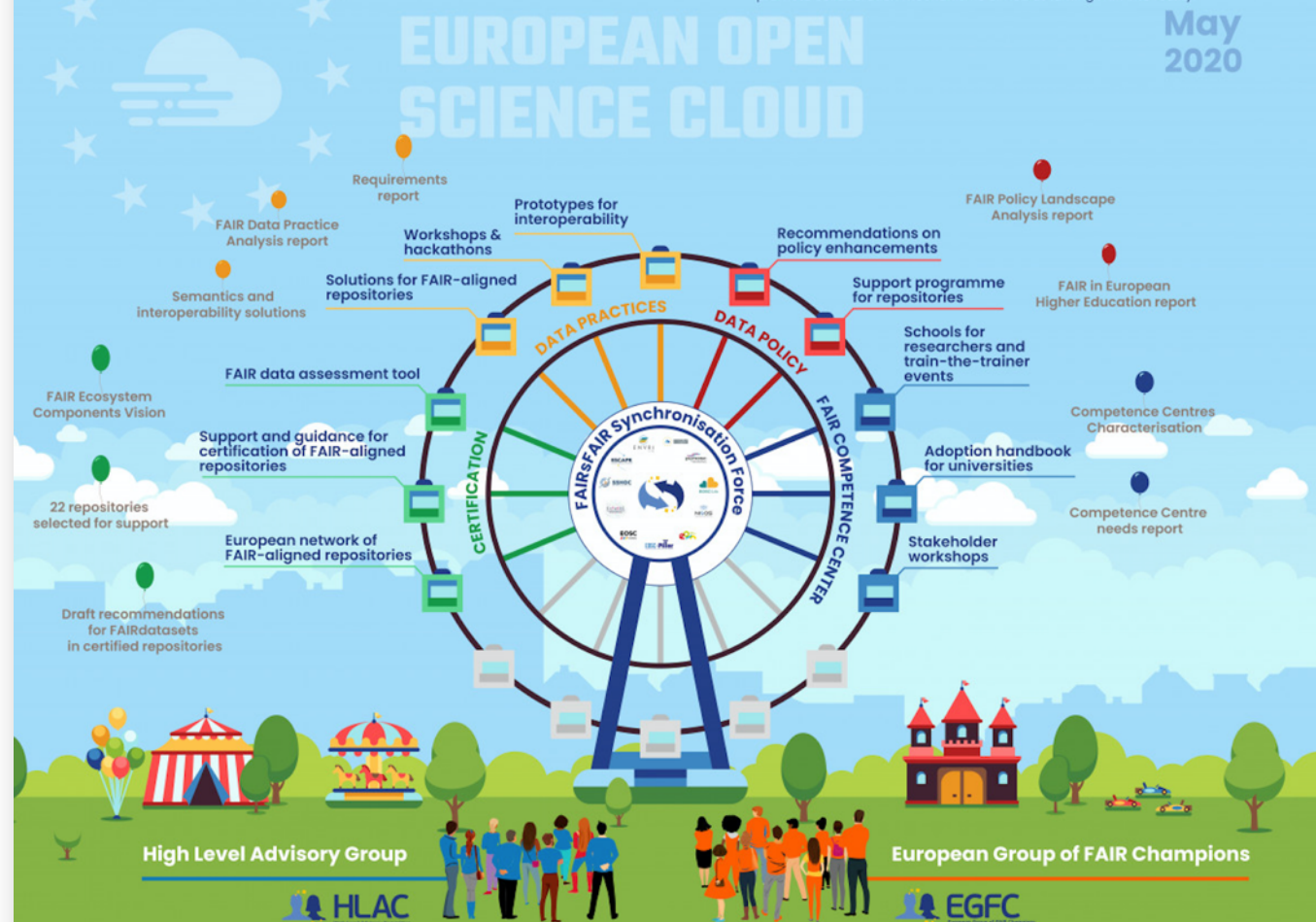
Embracing a FAIR Culture

FAIRsFAIR Synchronisation efforts in aligning with EU initiatives to foster Open Science in Europe.

A key challenge for FAIRsFAIR is to ensure project activities dovetail with work carried out by the EOSC Governance Working Groups, and feed into and complement the work being done by other projects in the research data and FAIR space.

For this reason FAIRsFAIR set up the Synchronisation Force, a team tasked with establishing a dialogue among the various projects and actors in both the EOSC and FAIR ecosystems, whose work touches on FAIR. Its mandate is to maximise coordination and minimise unnecessary overlap or duplication, facilitate synergies between project activities, and EOSC governance, and promote collaboration mechanisms aimed at turning FAIR into reality.

May
2020



JOIN OUR COMMUNITY!



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



www.fairstair.eu



@FAIRsFAIR_eu



/company/fairstair

Projects and initiatives



EOSC governance

(5a)



EOSC
secretariat.eu

Setup and management of the EOSC Secretariat supporting the EOSC Governance

Regional Nodes / Thematic Projects



FAIR (5C)




EOSC Governance Board




EOSC Executive Board




EOSC Working Groups

Landscape WG

Sustainability WG

Rules of Participation WG

Architecture WG

FAIR WG

Skills and Training WG



INFRAEOSC-5 Cross Project Collaboration Board (CPCB)

INFRAEOSC-5 Task Forces:

- Landscaping
- FAIR data and Infrastructures
- Services onboarding
- National policies and governance
- Training and skills
- Dissemination and events

FAIR WG Task Groups

- FAIR practice
- Interoperability
- PIDs
- Metrics and certification



EOSC Interest Groups

- Researcher engagement and use cases
- Service and research product catalogue
- Federating core
- Glossary

"Horizontal"



ESFRI Clusters



Other FAIR Initiatives




FAIRSFAR

Fostering Fair Data Practices in Europe



DATA PRACTICES

◆ Reports

- ◆ FAIR requirements for persistence and interoperability
- ◆ Guidelines for ontology design and vocabulary interoperability
- ◆ Basic framework for services enabling FAIR (including software)

◆ Solutions for interoperability and machine accessibility for FAIR-aligned repositories

- ◆ Prototype for interoperability of repositories
- ◆ Workshops and hackathons: Recommendations for FAIR Semantics and Semantics in FAIR



DATA POLICY

◆ Reports

- ◆ Recommendations on data policy and analysis of practice
- ◆ Integration of meta-data catalogues
- ◆ White paper on alignment and synchronisation around FAIR, Open Science and EOSC

◆ Support programme for repositories to reach FAIR compliance



CERTIFICATION

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



TRAINING, EDUCATION AND SUPPORT

◆ Reports

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

MAIN OUTPUTS

March 2019 - February 2022



FAIRSFAR "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

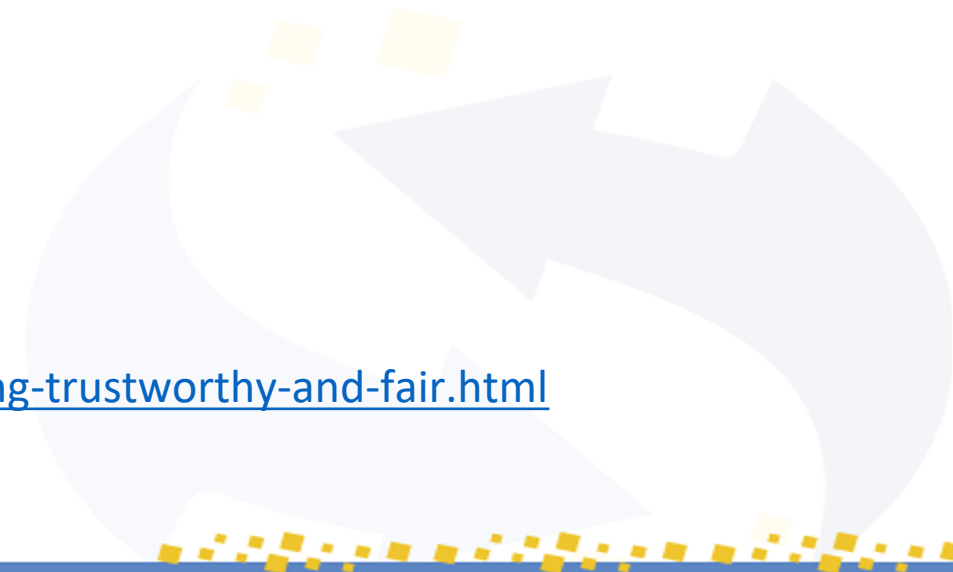
Data sharing a FAIRytale?




“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, CoreTrustSeal Board

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



The background image is the cover of a report titled 'The Data Harvest'. It features a golden wheat field in the foreground with a yellow combine harvester in the distance. The sky is blue with faint, stylized binary code (0s and 1s) and white lines suggesting data flow or connectivity. The text is overlaid on the left side of the image.

“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?”

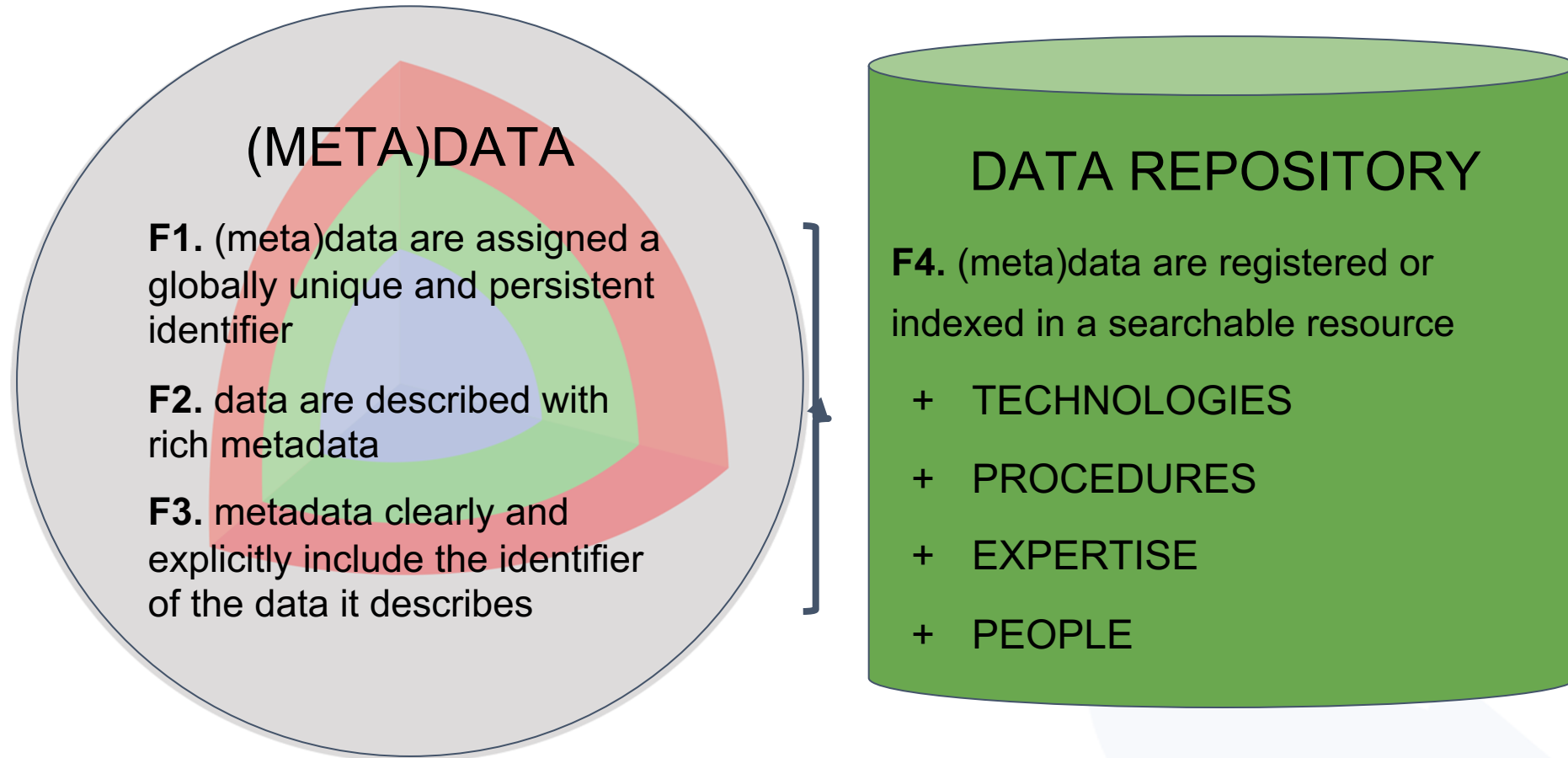
The Data Harvest:

How sharing research data can yield knowledge, jobs and growth

An RDA Europe Report

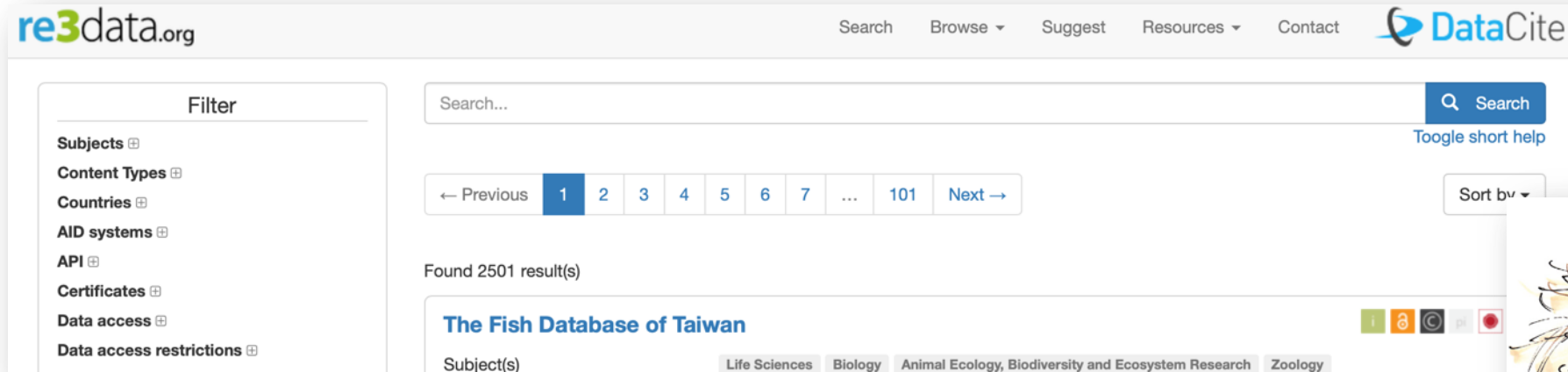
December 2014

FAIR data assessment: findable

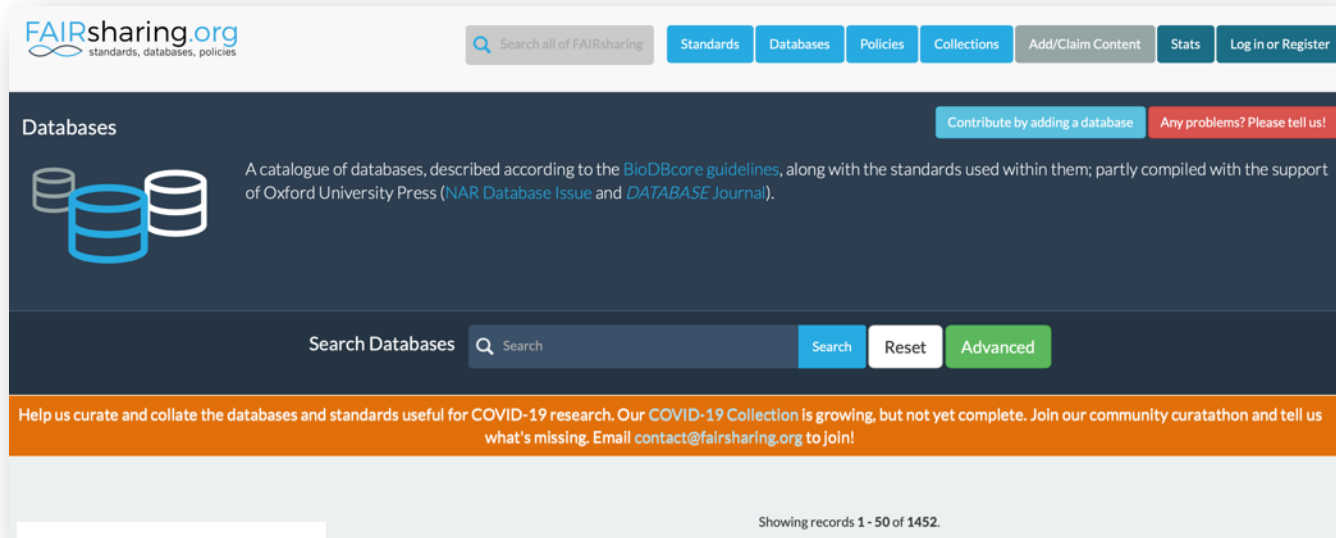


Where do you store your data?

www.re3data.org



The screenshot shows the re3data.org website interface. At the top, there's a navigation bar with links for Search, Browse, Suggest, Resources, and Contact, along with the DataCite logo. A search bar is prominently displayed. On the left, a 'Filter' sidebar lists categories like Subjects, Content Types, Countries, AID systems, API, Certificates, Data access, and Data access restrictions. The main content area shows search results for 'The Fish Database of Taiwan', indicating 2501 results. Below the title, there are tags for 'Life Sciences', 'Biology', 'Animal Ecology, Biodiversity and Ecosystem Research', and 'Zoology'. A pagination bar shows the current page is 1 of 101.



The screenshot displays the FAIRsharing.org 'Databases' page. The header includes the FAIRsharing.org logo and a search bar. Navigation tabs for Standards, Databases, Policies, Collections, Add/Claim Content, Stats, and Log in or Register are visible. The main section features a database icon and a description: 'A catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them; partly compiled with the support of Oxford University Press (NAR Database Issue and DATABASE Journal)'. Below this is a search bar with 'Search', 'Reset', and 'Advanced' buttons. An orange banner at the bottom encourages users to help curate databases for COVID-19 research. The footer indicates 'Showing records 1 - 50 of 1452'.



<https://fairsharing.org/databases/>

Trusting data repositories

- actions and attributes of the trustee (integrity, transparency, competence, predictability, guarantees, positive intentions)
- external acknowledgements:
 - reputation (researchers)
 - third party endorsements (funders, publishers)



CoreTrustSeal certification

- Community driven repository certification standard
- Developed under the umbrella of RDA
- 16 requirements (organizational infrastructure, digital object management, technology and security)
- Peer review, 3 year cycle, transparent processes
- Global uptake, discipline agnostic

<https://www.coretrustseal.org>



TDR to guarantee baseline data FAIRness

- Majority of CoreTrustSeal requirements (indirectly) refer to the FAIRness of the repository holdings
- Baseline of data FAIRness, but:
- Some data will be more FAIR than others!



TRUST principles to complement FAIR principles

www.nature.com/scientificdata

SCIENTIFIC DATA

110110
0111101
11011110
011101101

Check for updates

OPEN

COMMENT

The TRUST Principles for digital repositories

Dawei Lin^{1✉}, Jonathan Crabtree², Ingrid Dillo³, Robert R. Downs⁴, Rorie Edmunds⁵, David Giaretta⁶, Marisa De Giusti⁷, Hervé L'Hours⁸, Wim Hugo⁹, Reyna Jenkyns¹⁰, Varsha Khodiyar¹¹, Maryann E. Martone¹², Mustapha Mokrane³, Vivek Navale¹³, Jonathan Petters¹⁴, Barbara Sierman¹⁵, Dina V. Sokolova¹⁶, Martina Stockhause¹⁷ & John Westbrook¹⁸

As information and communication technology has become pervasive in our society, we are increasingly dependent on both digital data and repositories that provide access to and enable the use of such resources. Repositories must earn the trust of the communities they intend to serve and demonstrate that they are reliable and capable of appropriately

The TRUST principles offer guidance for the increase of the trustworthiness of digital repositories, especially of research data.

The principles are based various certification approaches CTS and ISO 163636 and translate the essence of these 'standards' into a set of principles that will be more appealing to an audience that is not familiar with digital preservation.

The principles are a means to facilitate communication with all stakeholders, including publishers.

Box 1 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.



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

Takeaway message



- Research data need to be shared 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 certification mechanisms, like CoreTrustSeal, help us to create trust in the research data infrastructure we need to safeguard the accessibility and assessability of our (FAIR) data for the future;
- Publishers represent an important stakeholder group in the promotion of data sharing;
- Publishers could support the FAIR and open sharing of data by:
 - endorsing the TRUST principles
 - pointing their authors to certified TDRs

Thank you for listening!



www.dans.knaw.nl

www.fairsfair.eu

ingrid.dillo@dans.knaw.nl

