# FAIR and TRUST a perfect couple

Ingrid Dillo, Deputy Director DANS UCLAC Webinar Series October 2020







Topics



Data Archiving and Networked Service

## DANS is about keeping data FAIR

Mission: enhancing scientific quality by facilitating reuse of 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

11111

DANS

National Centre of Expertise Research Data

Data Archiving and Networked Service:

## **DANS** Core Services



DANS

## **DANS** Core Services



Data Archiving and Networked Services



DANS Data sharing: the why

Data quality

FAIR data

Trust

**Repository certification** 

Data Archiving and Networked Service

'Just as the Netherlands is a knowledge society, science is a knowledge community. Both benefit from the free exchange of information. For that reason alone you have my fullest support in your quest for Open Science.'

Ingrid van Engelshoven, minister of Education, Culture and Science



free exchange of information for the good of science and society at large MAKING OPEN SCIENCE

A REALITY





Data Archiving and Networked Services

## **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



## Why data sharing is important

# Replication and validation of research outcomes (scientific integrity and transparency)





# nature International weekly journal of science

**Dutch universities** 

News

# The New York Times

Fraud Case Seen as a Red Flag for Psychology Research

By BENEDICT CAREY Published: November 2, 2011

A well-known psychologist in the Netherlands whose work has been published widely in professional journals falsified data and made up entire experiments, an investigating committee has found. Experts say the case exposes deep flaws in the way science is done in a field, psychology, that has only recently earned a fragile respectability.



Search a

# **SPIEGEL** ONLINE

#### Niederlande

Renommierter Psychologe gesteht Fälschungen

#### Report: Dutch 'Lord of the Data' Forged Dozens of Studies (UPDATE)

**Report finds massive fraud at** 

Investigation claims dozens of social-psychology papers

Breaking news and analysis from the world of science policy

by Gretchen Vogel on 31 October 2011, 7:05 PM | 34 Comments

contain faked data.



#### Public sector can combat fraud with data sharing

*Science* **Insider** 

Outsourcing is not the only thing to blame for procurement fraud, says Graham Kemp, and the public sector needs to view data less as a security risk but knowledge to be shared

PLosone

Home Browse Articles About For Readers For Authors and Reviewers

#### **RESEARCH ARTICLE**

How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data

## Why data sharing is important

# Re-use of data (efficiency, return on investment, standing on the shoulders of others)



Data Archiving and Networked Services



# 16 year study suggests air temperature is external trigger for heart attack

28 Aug 2017

#### Topic(s): Environmental and Cardiovascular Disease;

**Barcelona, Spain - 28 Aug 2017:** A 16 year study in more than 280 000 patients has suggested that air temperature is an external trigger for heart attack. The findings are presented today at ESC Congress. (1)

"There is seasonal variation in the occurrence of heart attack, with incidence declining in summer and peaking in winter," said first author Dr Moman A. Mohammad, from the Department of Cardiology at Lund University, Skane University Hospital, Lund, Sweden. "It is unclear whether this is due to colder temperatures or behavioural changes."

This nationwide, 16 year, observational study led by Prof David Erlinge from Lund University, is the largest to investigate the association between heart attack incidence and weather conditions such as air temperature, sunshine duration, precipitation, and air pressure.

Using the Swedish myocardial infarction registry (SWEDEHEART), all consecutive heart attacks treated at a coronary care unit between 1 January 1998 and 31 December 2013 were included in the study. The investigators studied the specific weather conditions during which heart attacks occurred using local meteorological data from hundreds of weather stations in the Swedish Meteorological and Hydrological Institute (SMHI).

https://www.escardio.org/The-ESC/Press-Office/Press-releases/16year-study-suggests-air-temperatureis-external-trigger-for-heart-attack







## ...but what about the researchers?







Data sharing: the why

### **Data quality**

FAIR data

Trust

**Repository certification** 

## Data quality



Trust is a central element in research.

The data re-user wants to know:

- Where do these data come from?
- How were they collected?
- What has happened with them along the way?

Quality - provenance



## Quality dimensions



- Scientific quality
- Fitness for use
- Technical quality



# Dimension 1: Scientific quality

#### **Principles:**

Honesty, scrupulousness, transparency, independence, responsibility.

#### Shared responsibility:

- the *individual* responsibility of the researcher;
- the *institutional* context within which the research is carried out;
- the informal *networks* that play a vital role within the research community.



ADVISORY REPORT



## Dimension 2: Fitness for use

- definition of data quality as "fitness for use"
- data quality judgement depending on data consumers

Wang, R. Y., & Strong, D. M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12(4),





## Dimension 3: Technical data quality



Based on UK Data Archive lifecycle: https://www.ukdataservice.ac.uk/managedata/lifecycle

- As a product, data have quality, resulting from the process by which data are generated.
- Managing and documenting data through all stages helps to build trust.







DANS Data sharing: the why Data quality

**FAIR data** 

Trust

**Repository certification** 

Data Archiving and Networked Service

## FAIR Data Principles (2014)

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



- F: Easy to find by both humans and machines based on metadata
- A: With well-defined use license and access conditions (Open Access if possible)
- I: Ready to be linked with other datasets
- R: Ready to be re-used for future research and to be processed further using computational methods and tools

## FAIR guiding principles (2016)

nature > scientific data > comment > article

#### 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 Ser - Show fewer authors

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

(1) An Addendum to this article was published on 19 March 2019

a naturerese	earch journal				
Search E-alert	Submit Login				
Download PDF	$\overline{1}$				
7071322CitationsAltmetric	Article metrics ≫				
Sections	References				
Abstract					
Comment					
Additional Information					
References					
Acknowledgements					
Author information					
Rights and permissions					
About this article					

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



## FAIR metrics

#### **Box 2** | The FAIR Guiding Principles

#### To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

#### To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

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

#### To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

12. (meta)data use vocabularies that follow FAIR principles

13. (meta)data include qualified references to other (meta)data

#### To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards



See: <u>http://datafairport.org/fair-principles-living-document-menu</u> and <u>https://www.force11.org/group/fairgroup/fairprinciples</u>





## Researchers and FAIR



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



## Researchers and FAIR

#### **Compliance with FAIR principles**

Of the participants who were familiar with FAIR, about a third said that their data management practices were very compliant with the principles. That proportion is similar across scientists at different stages of their career.

Very much Somewhat Neutral / Not very much



#### **High Awareness of FAIR Principles**



#### https://www.fairsfair.eu/fair-european-higher-education

#### Familiarity with FAIR principles

The majority of researchers surveyed as part of a recent study on open data had never heard of FAIR, regardless of their field. Of the 748 researchers that responded to this question, 144 said they were familiar with the principles. Circles are sized by number of respondents.





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

scontinuum • Findoble • Accessible MORE • Interoperatul **OPEN DATA** THAN • Beosable



## Responsible management of your data !?





## FAIR principles

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





## 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



## 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** ٠ and authorization procedure, where necessary
- **A2.** metadata are accessible, even when the **data** ٠ 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





#### DANS

Data sharing: the why

Data quality

FAIR data

Trust

**Repository certification** 

Data Archiving and Networked Services

"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

## Importance of sustainable data infrastructure

"36% of respondents have lost data on which they were working and there is, unsurprisingly, a high correlation between the vehicle for storing data and where it was lost - computer hard drives were the most common culprit here."

Science, Digital; Hahnel, Mark; Treadway, Jon; Fane, Briony; Kiley, Robert; Peters, Dale; et al. (2017): The State of Open Data Report 2017. figshare. Paper.

https://doi.org/10.6084/m9.figshare.5481187.v1





## Where to store your data? And whom should you trust?

Filter	Search
Subjects 🕀	
Content Types ⊕ Countries ⊕	$\leftarrow \text{Previous}  1  2  3  4  5  6  7  \dots  104  \text{Next} \rightarrow$
AID systems 🕀	
	Found 2580 result(s)
	HERE-WE HAVE TO FIND IT.



## A matter of TRUST: principles to **keep** data FAIR

•Discussion of a TRUST White Paper at two RDA plenaries

•Community consultation resulting in 200+ public comments



- •19 co-authors representing:
  - 4 continents, 8 countries
  - Diverse stakeholders: funders, publishers, librarians, standards experts, preservation specialists
  - Multiple domains: social sciences, biomedical sciences, geosciences, etc.



High level principles to facilitate stakeholder discussion and guide repositories

## 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.

**U**ser Focus To ensure that the data management norms and expectations of target user communities are met.

**S**ustainability To sustain services and preserve data holdings for the long-term.

Technology To provide infrastructure and capabilities to support secure, persistent, and reliable services.



Data Archiving and Networked Service.

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

## The TRUST principles

#### READ

The TRUST Principles for digital repositories https://doi.org/10.1038/s41597-020-0486-7

#### ENDORSE

https://www.rd-alliance.org/rda-community-effort-trustprinciples-digital-repositories

#### **CONTRIBUTE**

RDA Interest Group Certification: TRUST principles to be discussed at RDA Plenary 16







#### DANS

Data sharing: the why

Data quality

FAIR data

Trust

**Repository certification** 

Data Archiving and Networked Service

## Trusting digital repositories

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





## CoreTrustSeal

- Community driven repository certification
- Developed under the umbrella of RDA
- 16 requirements, reflecting the characteristics of TRUSTworthy Data Repositories (TDRs)
- Minimal (core) standard
- Peer review, 3 year cycle, transparent processes
- Global uptake, discipline agnostic



https://www.coretrustseal.org





- Gives data producers assurance that data are preserved and remain reusable (i.e. FAIR) in the future;
- Provides **funding bodies** confidence that investments are maximized;
- Enables **data consumers** to choose the repositories where data are held;
- Supports **repositories** to improve their processes.





## CoreTrustSeal requirements

#### Categories:

- Background information (R0)
- Organizational infrastructure (R1-6)
- Digital object management (R7-14)
- Technology and security (R15-16)
- Applicant feedback

#### DOI 10.5281/zenodo.168411

25/08/2015

Common Requirements/V2.1



WORLD DATA SYSTEM

#### DSA–WDS Partnership Working Group Catalogue of Common Requirements

#### Introduction

#### Importance of Certification

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that the data they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.

#### **Basic Certification and its Benefits**

Nowadays certification standards are available at different levels, from a basic level to extended and formal levels. Even at the basic level, certification offers many benefits to a repository and its stakeholders.





**Self assessment** based on 16 Requirements (written responses + URLs of documented public evidence + compliance level)

**Peer review** by two expert and independent reviewers under the responsibility of the CoreTrustSeal Standards and Certification Board

- Online tool
- Administrative fee of 1,000 euro
- Successful applications are made publicly available
- Certification valid 3 years

Library of public applications; all are certified and so can be considered exemplars.



## 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



## CoreTrustSeal uptake





## Current situation in LAC countries

- 32 repositories registered in the Re3Data registry
- No CTS certified trustworthy digital repositories yet







- 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
- We need more assessment mechanisms for in the FAIR ecosystem





## European FAIRsFAIR project

Fostering FAIR Data Practices in Europe

FAIRSFAIR

About Us Partners News Events Videos Contact Us

FAIRsFAIR aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle with emphasis on fostering FAIR data culture and the uptake of good practices in making data FAIR.

#### COMPETENCE FRAMEWORK

- A tested Framework for franchising data science schools with model courses and curricula
- "Train the trainer" data science schools
   FAIR Competence Adoption booklet
- Mapping existing FAIR data training offerings across education institutions

#### **REGISTRY for FAIR**

- Registry for FAIR compliant repositories
   Technical solutions for interoperability
- Technical solutions for interoperability requirements
- Training, support & guidance

### TOOLSET and REPOSITORIES • Toolsets on certified repositories to researchers • Core level certified repositories

- Badges for end-users
- Capability maturity model towards FAIR
- certification
  A Network of 50+ trusted digital repositories

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;

- 10 million euro
- 36 months
- Starting date: March 1 2019
- 22 partners from 8 European countries

https://www.fairsfair.eu/

Data Archiving and Networked Services

## FAIRsFAIR results

Wpl	Project Management and Sustainability
Wp2	FAIR Practices: Semantics, Interoperability, and Services
Wp3	FAIR Policy and Practice
Wp4	FAIR Certification
Wp5	Engagement, Communication and Uptake
Wp6	FAIR Competence Centre
Wp7	FAIR Data Science and Professionalisation

**FAIR** Data Object Assessment Metrics

P All versions	Found 63 results.	< 1 2 3 4 >	Sort b
Access Right			asc.
🗆 Open (63)	January 31, 2020 (v01.00) Other Open Access	1	View
	FAIRsFAIR Project Comments on	າ Data Repository Selection- Criteria That Mat	ter
	🚯 Hervé L'Hours; 💿 Ilona von Stein; 💿 Mu	istapha Mokrane; 💿 Joy Davidson; 💿 Robert Ulrich; 💿 Sara	ila Wimalaratne;
File Type	Criteria, characteristics and requirements r project. We provide a general response to t	related to data objects and data services, including reposito the paper, identify alignment with ongoing FAIRsFAIR work.	ries, are at the heart of the FAIRsFAIF and make some broad comments or
Pdf (54)	the proposed crite		
Pptx (5)	Uploaded on January 31, 2020		
🗆 Xlsx (5)			
Docx (3)	June 20, 2019 (v1) Presentation Open Access	I	View
🗆 Mp4 (3)	CODATA-RDA Schools		
Csv (2)	(i) Shananan, Hugh; A presentation made at the kick off meeting.	ag of FAIDoFAID	
🗆 Xls (1)	Liploaded on June 20, 2019	IG OF FAIRSFAIR.	
🗆 Zip (1)	opioudea on ourie 20, 2015		
	May 29, 2020 (1.0) Project deliverable Open Acc	cess	16
Kanana ala	D5.2 PAN-EUROPEAN UPTAKE IN	NTERIM REPORT	view
Keywords	o Sara Pittonet; 📀 Rita Meneses; Tracey B	Biller; 💿 Frans Huigen; 💿 Joy Davidson; 💿 Lennart Stoy; 💿	Jessica Parland; 💿 Yann Le Franc; 🌔
🗆 FAIR (15)	Anusuriya Devaraju;		
GFAIRsFAIR (14)	FAIRSFAIR aims to supply practical solution fostering FAIR data culture and the uptake	ns for the implementation of FAIR data principles througho of good practices in making data FAIR. To achieve this goa	ut the research data life cycle by I, it is crucial that targeted
FAIR Data (10)	stakeholders - both end users of		
	Uploaded on June 15, 2020		
	June 5, 2020 (1.0) Dataset Open Access		View
	Data for D7.1 FAIR in European Higher Education		

Uploaded on June 5, 2020

Publication (37) Presentation (13) Dataset (6) Other (3)

Uvideo (2) Poster (1)

Software (1)

💿 F-UJI

January 28, 2020 (1.0_BRAFT) Project deliverable Open Access View
D5.3 Report on the First Synchronisation Force Workshop
💿 Coen, Gerard; 💿 Mokrane, Mustapha; 💿 Pittonet, Sara; 🔞 Hodson, Simon; 💿 van Kessel-Hagesteijn, Renee;
This is the report of the FAIRsFAIR Synchronisation Force workshop organised in Budapest, Hungary on 25th November 2019 in conjunction with the EOSC Symposium. It highlights the points discussed by stakeholders from the FAIRsFAIR project, EOSC Working Groups, and other associated EOSC representative
Uploaded on February 10, 2020
November 21, 2019 (v1) Dataset Open Access View
FAIRsFAIR Characterisation of Competence Centres
💿 Herterich, Patricia; 💿 Davidson, Joy; 💿 Grootveld, Marjan; 💿 Whyte, Angus; 💿 Molloy, Laura; 💿 Matthews, Briar; 💿 Kayumbi Kabeya, Gabin;

Stay tuned:

Software & Tools \_\_\_\_

**FAIR** Aware

https://www.fairsfair.eu/user/register

#### https://zenodo.org/search?page=1&size=20&q=fairsfair



https://www.rd-alliance.org/plenaries/rda-16thplenary-meeting-costa-rica

#### **RDA 16th Plenary - Pathways**

Home » VP16; P16; Plenary 16; Pathways » RDA 16th Plenary - Pathways

🖻 20 October 2020 🛛 177 reads 🛛 🖪 Facebook 🗳 Twitter

#### **Plenary 16 Information**

#### Welcome to the RDA Plenary Pathways for P16!

#### Find your breakout session by theme.

The Plenary Pathways are the result of a joint effort between TAB and OA to provide a guide through the rich diversity of topics that the plenary presents. As such, they are meant to enhance the plenary experience rather than constrain, to highlight convergences, and to propose divergences you may not have considered. We hope you find them useful.

Please find the Pathways below:

- COVID-19 / Infectious Diseases
- ➤ The Data Life Cycle, Versioning, Provenance, and Reuse
- Research Software
- Data Management Practices
- Skills and Human Capacity Development
- The FAIR Agenda

## Main takeways

- 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;
- Certification mechanisms, like CoreTrustSeal for digital repositories, help us to create **TRUST** in the research data infrastructure we need in order to safeguard the accessibility and assessibility of our (FAIR) data for the future.







www.dans.knaw.nl

Data Archiving and Networked Servic