UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

https://www.flickr.com/photos/infocux/8450

Open Data and FAIR Data

Presented at the Open Access Workshop, held 29 May 2019, Merensky Library, University of Pretoria, Pretoria, South Africa

Johann van Wyk

What is Open Data?

"Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike". In other words without any restrictions (Open Data Handbook, n.d.)

Open means anyone can freely access, use, modify, and share data for any purpose (Open Knowledge Foundation, n.d.)

Key Features of Open Data

Availability and access:

the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

Reuse and redistribution:

the data must be provided under terms that permit reuse and redistribution, including the intermixing with other datasets. The data must be machine-readable.

Universal participation:

everyone must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

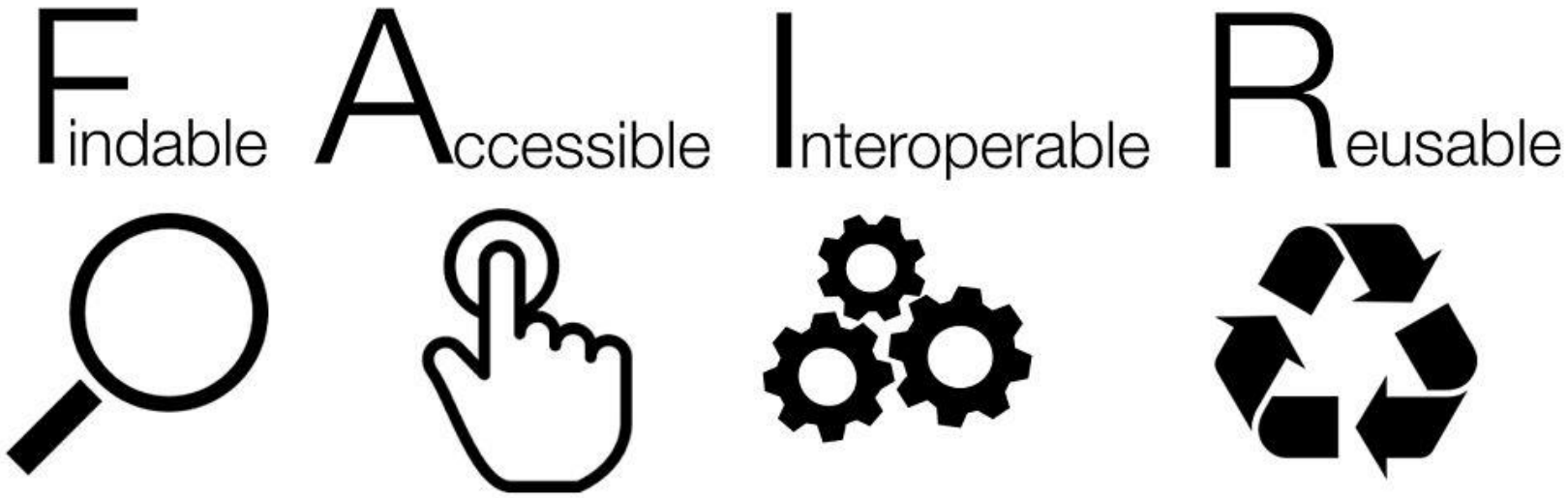
Reasons why people do not want to make their data open

Terrorists and/or criminals will use it	Its too complicated	People may misinterpret the data	Its not very interesting
I don't mind, but someone else might	I might want to use it in another paper/article	People will contact me to ask about the research	National security may be compromised
It's too big	Others will see that my data is embarrassingly bad	It's not useful to anyone else	It's not a priority and I do not have the time
I don't know how	I'm not sure I own the data	Someone might steal/plagiarise the data and not credit me	My funder does not require it

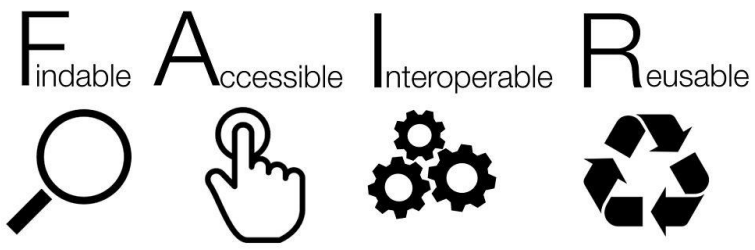
Reasons for Open Data

- Transparency (Open Knowledge International, n.d.)
- Good scientific practice depends on communicating the evidence (Hodgson, 2018)
- It will enhance the visibility of one's research and increase one's citations (McArthur, 2015) (Emory Libraries and Information Technology, 2019)
- It benefits science – science is based on building on, reusing and openly criticizing scientific knowledge (McArthur, 2015)
- Open data can foster innovation and accelerate scientific discovery through the reuse of data (Hodgson, 2018)
- Allows others to reproduce experiments and verify results
- Reduces inefficiencies, including duplication of research (CODATA, 2015)
- Reduces academic fraud (Emory Libraries and Information Technology, 2019)
- Its for the benefit of society (McArthur, 2015)
- Most research is publicly funded (McArthur, 2015)
- Ensure compliance with funder and publisher mandates. Many funders and publishers are now requiring data to be made openly available (McArthur, 2015) (Emory Libraries and Information Technology, 2019)

FAIR Data



What is FAIR?



FAIR is a set of principles that ensure that data are shared in a manner that enables and enhances re-use by humans and machines

Where did the idea of FAIR come from?

Emerged from a workshop held at the Lorentz Centre in Leiden, Netherlands in 2014

Come from life sciences, but is intended for all types of data

Issued by FORCE11 (community of scholars, librarians, archivists, publishers and research funders)

https://www.force11.org/group/fairgroup/fairprinciples

Recommended by Association of European Research Libraries (LIBER)

What does FAIR mean?

Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

Reusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

From: CESSDA Data Management Expert Guide

Benefits of FAIR Data Principles

- Gain maximum potential from data assets
- Increase the visibility and citations of research
- Improve the reproducibility and reliability of research
- Stay aligned with international standards and approaches
- Attract new partnerships with researchers, business, policy makers and broader communities
- Enable new research questions to be answered
- Use new innovative research approaches and tools
- Achieve maximum impact from research

(CCAFS, n.d.)

https://doi.org/10.1038/sdata.2016.18

FAIR Principles

Findable

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

Interoperable

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

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified

references to other (meta)data.

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.

Reusable

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

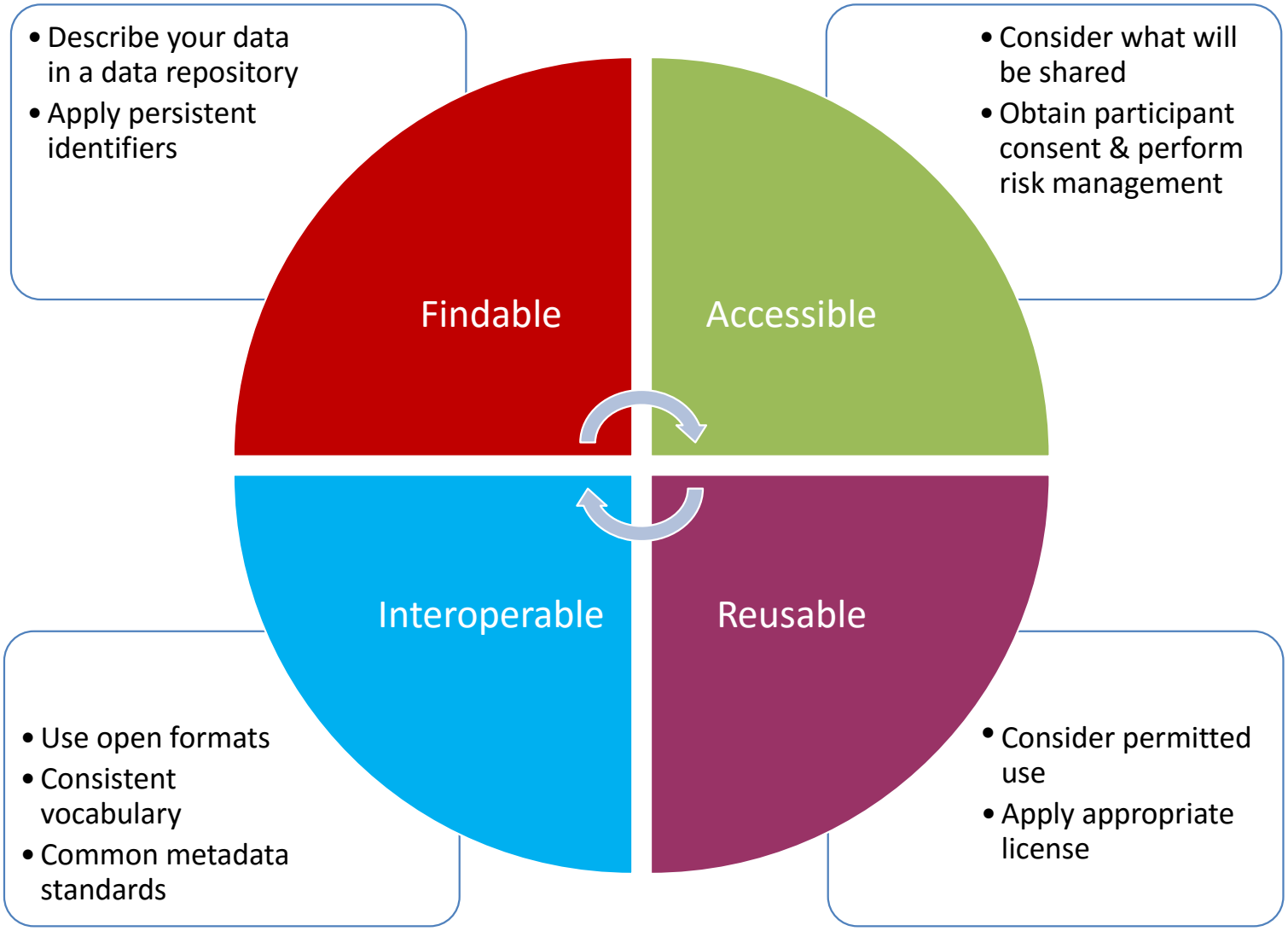
R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

FAIR Principles

https://www.slideshare.net/lshtm/preparing-data-for-sharing-the-fair-principles

FAIR Data



FAIR Self-assessment Tool

The Australian ARDC (an initiative between ANDS, Nectar and RDS) developed a FAIR self-assessment tool that enables you to assess the FAIRness of a dataset, and also helps you to determine how to enhance a dataset's FAIRness.

https://www.ands-nectar-rds.org.au/fair-tool

Does FAIR Data = Open Data?

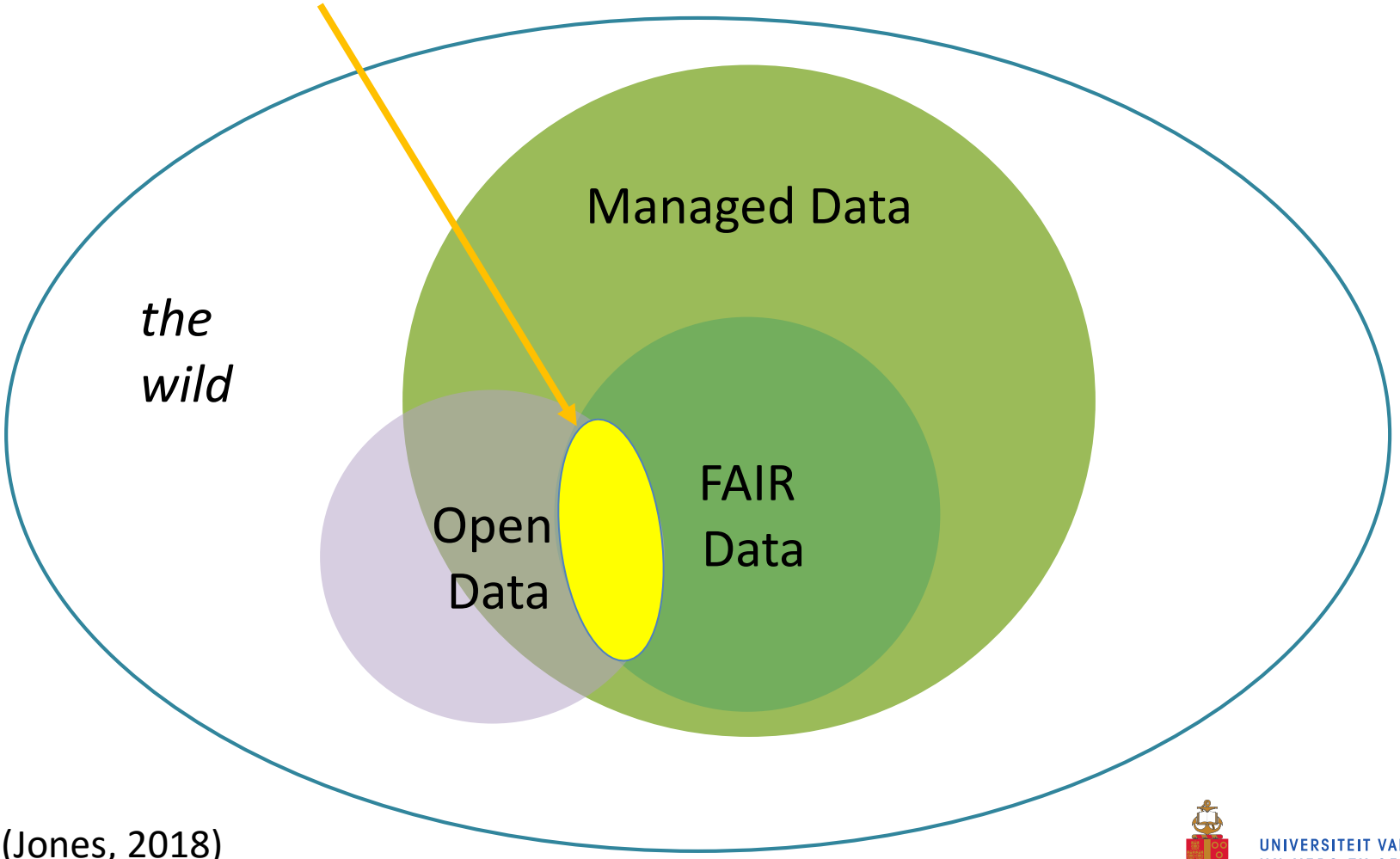
- FAIR data does not have to be Open
- Data can be FAIR or Open, both or neither (Hodgson, 2018)
- Data can be shared under restrictions & still be FAIR
- Making data FAIR ensures it can be found, understood and reused
- Even though Open Data and FAIR Data are different, they can be overlapping concepts; FAIR data does not automatically imply that it needs to be accessible - there can be limitations to access, for example, for sensitive data.
- Accessibility of FAIR data means “how-to-access”, and is defined in a human- and machine-readable way. (The AIMS team, 2018)

Does FAIR Data = Open Data?

- FAIR data can be accessed by appropriate people, at an appropriate time, and in an appropriate way, depending on the purpose of the data.
- In other words, data can be FAIR when it is private, when it is accessible by a specific group of people or when it is accessible by everyone.
- Open Data though are available for all to access, use and share, without licenses, copyright, or patents. Only subjected to attribution/share alike licenses.
- Greatest potential for reuse and value comes when data are both FAIR and Open (Hodgson, 2018)

All Research Data

Increase that which is FAIR and Open



(Jones, 2018)

International Initiatives

- Enabling FAIR Data Project - http://www.copdess.org/enabling-fair-data-project/
- RDA FAIR Data Maturity Model Working Group
  https://www.rd-alliance.org/groups/fair-data-maturity-model-wg
- FAIRsFAIR https://www.fairsfair.eu/
- FOSTER Open Science Toolkit
  https://www.fosteropenscience.eu/node/2556
- GOFAIR https://www.go-fair.org/
- EOSC-hub https://www.eosc-hub.eu/
- OpenAIRE https://www.openaire.eu/
- Open Data Toolkit http://aims.fao.org/activity/blog/implement-effective-open-data-keep-it-alive-open-data-toolkit
- Search Engine for Open Data – Google Dataset Search (Beta)
  https://toolbox.google.com/datasetsearch

Thank you

References

- AUSTRALIAN NATIONAL DATA SERVICE (ANDS). n.d. ***FAIR resources graphic.*** Caulfield East, VIC, Australia: ANDS. [Online] available at https://www.ands.org.au/working-with-data/fairdata/training (Accessed 29 April 2019).

- CCAFS. n.d. ***Open access and FAIR principles.*** Wageningen, Netherlands: CCAFS Program Management Unit. [Online] available at https://ccafs.cgiar.org/open-access-and-fair-principles#.XNE3l44zaUk (Access 7 May 2019)

- CESSDA Training Working Group. 2017- 2018. ***CESSDA Data Management Expert Guide.*** Bergen, Norway: CESSDA ERIC. Retrieved from https://www.cessda.eu/DMGuide (Accessed 29 April 2019).

- CODATA. 2015. ***Report on the value of open data sharing.*** Paris, France: CODATA. [Online] available at https://zenodo.org/record/33830#.XOPTtsgzaUk (Accessed on 21 May 2019)

References

- EMORY UNIVERSITY LIBRARIES AND INFORMATION TECHNOLOGY. 2019. ***Benefits of open data to researchers.*** Atlanta, GA: Scholarly Communications Office, Emory University Libraries and Information Technology. [Online] available at https://sco.library.emory.edu/research-data-management/open-data/benefits-research.html (Accessed 29 April 2019).

- GOFAIR. n.d. ***What is the difference between FAIR data and Open data if there is one?*** Paris; Hamburg; Leiden: GOFAIR. [Online] available at https://www.go-fair.org/faq/ask-question-difference-fair-data-open-data/ (Accessed 29 April 2019).

- HODGSON, S. 2018. ***How can we build capacity for an open science and FAIR data ecosystem.*** Paper delivered at the Open Science Conference, Brussels, Belgium, 21 November 2018. [Online] available at https://rio.jrc.ec.europa.eu/en/file/12570/download?token=aV8Js7_x (Accessed 21 May 2019).

References

- JONES, S. 2018. ***Open data, FAIR data and RDM: the ugly duckling.*** Presented at the Open Science Conference , Berlin, Germany, 13-14 March 2018. [Online] available at https://www.slideshare.net/sjDCC/open-fair-data-and-rdm (Accessed 21 May 2019)

- MACARTHUR, J. 2015, ***Four reasons to do Open Data.*** Presented at RECODE Early Career Researcher Workshop held 14 May 2015 at University of Sheffield, Sheffield, UK. [Online available at https://www.slideshare.net/RightToResearch/4-reasons-to-do-open-data (Accessed 29 April 2019).

- MARLON, E. 2006***. Bingo.*** [Online] as cited by MacArthur, J. 2015, Four reasons to do Open Data. Presented at RECODE Early Career Researcher Workshop held 14 May 2015 at University of Sheffield, Sheffield, UK. [Online available at https://www.slideshare.net/RightToResearch/4-reasons-to-do-open-data (Accessed 29 April 2019).

- MURRAY-RUST, P.; NEYLON, C.; POLLOCK, R. & WILBANKS, JOHN. 2010. ***Panton Principles, Principles for open data in science.*** [Online] available at: https://pantonprinciples.org/ (Accessed 9 May 2019).

References

- OPEN KNOWLEDGE FOUNDATION. n.d. ***The open definition.*** London, UK: Open Knowledge Foundation. [Online available at http://opendefinition.org/ (Accessed 9 May 2019).

- OPEN KNOWLEDGE INTERNATIONAL. n.d. ***Open data handbook***, London: Open Knowledge International. [Online] available at http://opendatahandbook.org/guide/en/what-is-open-data/ (Accessed 29 April 2019).

- OPEN KNOWLEDGE INTERNATIONAL. n.d. ***What is open?*** London: Open Knowledge International. [Online] available at https://okfn.org/opendata/ (Accessed 29 April 2019).