

Making data FAIR - why and how?

Maggie Hellström

ICOS Carbon Portal & Lund university



LECCE
1-5 JUL
2019

INTERNATIONAL SUMMER SCHOOL
FOR ENVIRONMENTAL &
EARTH SCIENCE INFRASTRUCTURES



A lot of pressure on scientists...

WE WANT



YOUR DATA!

Image:imgflip.com

- Open Science = Open Access + Open Data
- Public-funded data should be shared
- Data is the new oil: data-driven innovation
- Data must be FAIR!
- The “reproducibility problem”
- Prove that your results are useful (and used)!
- Write a Data Management Plan – or else!

Surfing the wave – or drowning?



Image:unknown

- Big Data “tsunami”
- Not just research data – many different sources
- Not enough metadata to make sense of it all
- Fragmentation of information
- Too many standards & formats
- “Wrangling data” instead of “doing science”
- Increasing data & metadata losses

How can we become better at locating, accessing, interpreting and making use of (other people’s) data ????

FAIR principles to the rescue



- FAIR stands for Findable, Accessible, Interoperable, Reusable
- not a standard, but a set of 15 guiding principles
- encourages comprehensive & sustainable data management
- involves “everyone”: producers, stewards and users
- was coined in 2014 (see <https://www.force11.org/fairprinciples>)
- is increasingly called for by funders & policy makers

F for Findable

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



FORCE11, 2014 <https://www.force11.org/fairprinciples>

A for Accessible

- **A1.** (meta)data are **retrievable by their identifier** using a standardised 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**



FORCE11, 2014 <https://www.force11.org/fairprinciples>

I for 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



FORCE11, 2014 <https://www.force11.org/fairprinciples>

R for Reusable (and Reproducible)

- **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



FORCE11, 2014 <https://www.force11.org/fairprinciples>

What FAIR isn't

- FAIR is *not* a standard
- FAIR is *not equal to* 'Open' or 'Free'
 - Data are often Open but not FAIR
 - Data can be Closed, yet perfectly FAIR
- FAIR is no **guarantee** for successful (open) science



M. Wilkinson et al. (2018): *Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud*, <http://doi.org/10.3233/ISU-170824> and GO-FAIR (2018): *FAIR Data Stewardship Awareness Course*



What FAIR shouldn't be...

- applied only for research outputs
- just for computers
- confused with quality
- the main scoring criterion



M. Wilkinson et al. (2018): *Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud*, <http://doi.org/10.3233/ISU-170824> and GO-FAIR (2018): *FAIR Data Stewardship Awareness Course*



FAIR stakeholders

- “Society”
- Funders
- Research infrastructures
- Scientists
- Data stewards
- Librarians & data support units
- Repositories
- e-service providers

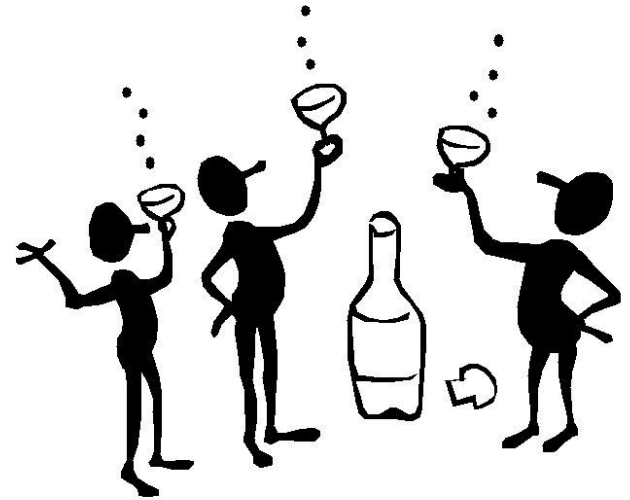


Image:123RF.com

What's in it for scientists?

Making your data “FAIR enough” gives you better control of what happens to your data, and

- helps make your data sustainable
- ensures your data can be found by others
- makes collection of metadata easier
- guarantees data can be cited
- facilitates data usage statistics
- simplifies reporting to funders & streamlining cost estimates
- makes data management plans easier to write



How can data become FAIR? (1/2)



- Make a **plan** for the data **before** you start a project!
- Collect detailed descriptive information (= **metadata**) throughout
- Use **standards** and **formats common** to your discipline
- Store the data in a **trusted & sustainable** repository or data center
- See to that the data gets **persistent identifiers** (DOIs)

How can data become FAIR? (2/2)

- Apply a suitable **usage license**
- Provide end users with information on **“intended use”**
- Make the data **“as open as possible, as closed as necessary”**
- Ensure that **metadata remain available** even if the data cannot be accessed any more
- FAIRness needs to be applied **where it makes sense**



Principles in principle



Image © 2016 by Brian McCarthy,
<https://pixels.com/featured/groucho-marx-brian-mccarthy.html>

“Those are my principles, and if you don't like them... well, I have others.” – *Groucho Marx* *)

*) Actually, to be FAIR it is not at all clear that Groucho Marx ever said this! According to the [Quote Investigator web site](#), the earliest recorded usage was in a New Zealand newspaper in 1873 – long before Groucho's illustrious career in comedy even started...

Learn more!

- “The FAIR data principles”, FORCE11 (2014), <https://www.force11.org/group/fairgroup/fairprinciples>
- GO-FAIR’s materials, including examples, about the principles: <https://www.go-fair.org/fair-principles/>
- “Turning FAIR into reality”, S. Jones ed. (2018), Report from the European Commission’s high level expert group on FAIR data, <http://doi.org/10.2777/1524>
- “Data stewardship for open science – Implementing FAIR principles”, B. Mons (2018), CRC Press, 244 pp. ISBN 9781498753180 (ebook), <https://doi.org/10.1201/9781315380711>
- “Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud”, B. Mons et al. (2017), <http://doi.org/10.3233/ISU-170824>
- “A design framework and exemplar metrics for FAIRness”, M. Wilkinson et al. (2017), bioRxiv preprint, <http://dx.doi.org/10.1101/225490>