



Data  
Schools

# RDM & Open Science: further steps & recent developments

*S. Venkataraman, OpenAIRE*



# Take stock!

- How much of what you learnt are you actively following?
- Which parts are most useful to you?
- Is there anything that you feel could be changed to better fit your needs?
- Is there anything you actively decided not to do or use – why not?
- Are there funder and/or host institution mandates on RDM?
- (check the [African Open Science Platform](#))

# Latest developments in FAIR

- EC [report](#) and recent [publication](#) on implementation considerations
- dedicated organisation established – [GO FAIR](#)
- large European (multi-million Euro) funded projects devoted especially to FAIR, e.g. FAIRsFAIR, FAIRplus, FAIR4Health
- see also this [FAIR in practice reference list](#) (includes examples from outside Europe!)

# ...and the UN SDGs

- In context of SRDS and its objectives
- UN's sustainable development goals (SDGs) – how can data science play a part?
- Database of SDG indicators can also be found



How data science and analytics can contribute to sustainable development



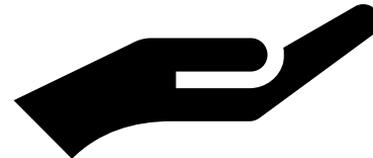
[www.unglobalpulse.org](http://www.unglobalpulse.org)  
©UNGlobalPulse 2017

- 1 NO POVERTY**  
Spending patterns on mobile phone services can provide proxy indicators of income levels
- 2 AFFORDABLE AND CLEAN ENERGY**  
Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods
- 3 GOOD HEALTH AND WELL-BEING**  
Mapping the movement of mobile phone users can help predict the spread of infectious diseases
- 4 QUALITY EDUCATION**  
Citizen reporting can reveal reasons for student drop-out rates
- 5 GENDER EQUALITY**  
Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women
- 6 CLEAN WATER AND SANITATION**  
Sensors connected to water pumps can track access to clean water
- 7 AFFORDABLE AND CLEAN ENERGY**  
Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods
- 8 DECENT WORK AND ECONOMIC GROWTH**  
Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP
- 9 INDUSTRY, INNOVATION AND INFRASTRUCTURE**  
Data from GPS devices can be used for traffic control and to improve public transport
- 10 REDUCED INEQUALITY**  
Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response
- 11 SUSTAINABLE CITIES AND COMMUNITIES**  
Satellite remote sensing can track encroachment on public land or spaces such as parks and forests
- 12 RESPONSIBLE CONSUMPTION AND PRODUCTION**  
Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products
- 13 CLIMATE ACTION**  
Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation
- 14 LIFE BELOW WATER**  
Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities
- 15 LIFE ON LAND**  
Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze
- 16 PEACE, JUSTICE AND STRONG INSTITUTIONS**  
Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights
- 17 PARTNERSHIPS FOR THE GOALS**  
Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today's hyper-connected world



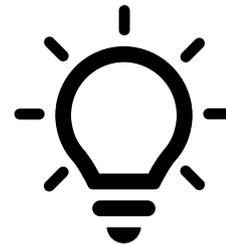
# Building institutional support

- There is a growing trend in institutions looking to provide the necessary tools for researchers
- Increase **ownership**
- Decrease use of third party (commercial) solutions
- Try the research infrastructure self evaluation ([RISE](#)) tool yourself to see how your institution fares



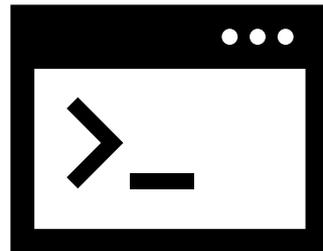
# Sensitive data

- Increasing interest in how to align these data with wider research data
- New, specific rules need to be developed
- Doesn't only mean clinical data – geospatial, IPR, etc
- Some examples: [Reproducible Health Data Services](#), [Raising FAIRness in health data and health research performing organisations \(HRPOs\)](#)



# Software

- Growing movement to apply FAIR to software and code
- Still treated as a “data” object
- [CURE-FAIR](#)
- See also Lamprecht, Anna-Lena et al. ‘Towards FAIR Principles for Research Software’. Data Science, vol. 3, no. 1, pp. 37-59, 2020. [DOI: 10.3233/DS-190026](https://doi.org/10.3233/DS-190026)
- See also [RSEs](#)





# Data Schools

[s.venkataraman@openaire.eu](mailto:s.venkataraman@openaire.eu)

[@OpenAIRE\\_eu](https://www.openaire.eu)

<https://www.openaire.eu/>



OpenAIRE | NEXUS



# Homework

Imagine you are a biologist who is doing microscopy experiments imaging tissue specimens. The data captured by the imaging is 100s of GB in size and is then cleaned and analysed to produce derivatives of the original captured data. Some of these derivatives may eventually be published. In preparation for publication, the data will also be segmented and annotated using standard ontologies. Documentation will also include metadata standards that will sufficiently describe the experimental procedure to allow reproducibility. Publication of the data is mandatory due to funder policy and must be deposited in a repository within 3 years of data production and must use an open licence without restrictions on reuse.

Now...please split into groups and see if you can answer the following questions using the tools and guidelines that have been described:

- What **file format(s)** should data be captured/preserved in?
- Which **metadata standard(s)** should be used?
- What **ontology(ies)** should be used?
- Which **licence(s)** should be used?
- Which **repository** would be the best fit for these data?
- Do you foresee any problems with the data?
- (Hint: not all the questions can be answered definitively! – but why not?)

# Homework (optional)

Try the FAIRsFAIR tool [FAIR Aware](#). See what your level of FAIR compliance is and your level of understanding. This is an optional exercise, if you have time, and can be discussed at the F2F sessions if necessary.