Qualitative data sharing

Agata Bochynska, PhD
agata.bochynska@ub.uio.no
Open Research group, University of Oslo Library

QualiFAIR breakfast seminar 28.03.2023

“**Open Science** has the potential of making the scientific process more **transparent**, **inclusive** and **democratic**. It is (…) a true game changer in bridging the science, technology and innovation gaps and fulfilling the **human right to science**.”

UNESCO Recommendation on Open Science

https://youtu.be/I3Wkvx_ZaFo
https://www.unesco.org/en/natural-sciences/open-science

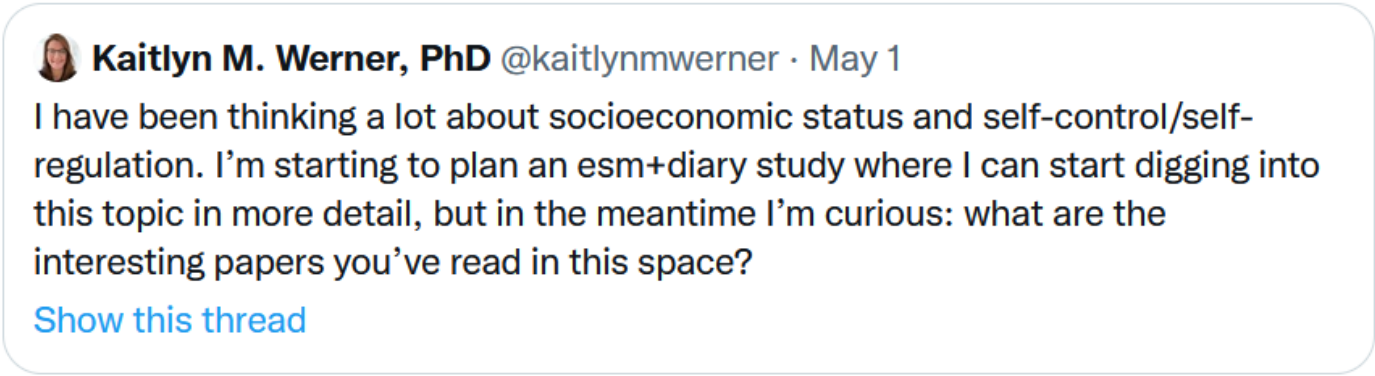
Local Requirements

Research data at the University of Oslo shall:

be **made openly available** for further usage
be made available at an **early stage**
have a **data management plan**
have **metadata** and be **documented**
must be securely **archived**
have **licenses** for access, reuse and redistribution
made **freely** available
(but the actual distribution cost should be covered)

Source: https://www.uio.no/english/for-employees/support/research/research-data-management/policies-and-guidelines/

Why open data?



Kaitlyn M. Werner, PhD @kaitlynmwerner · May 1

I have been thinking a lot about socioeconomic status and self-control/self-regulation. I'm starting to plan an esm+diary study where I can start digging into this topic in more detail, but in the meantime I'm curious: what are the interesting papers you've read in this space?

Show this thread

Kaitlyn M. Werner, PhD
@kaitlynmwerner

Open science truly is beautiful. Someone recommended a paper w. open data relevant to this question. Within minutes I was able to analyze my question because the data/code was so beautifully and efficiently organized -- the best I've seen! Major props to @russpoldrack and team.

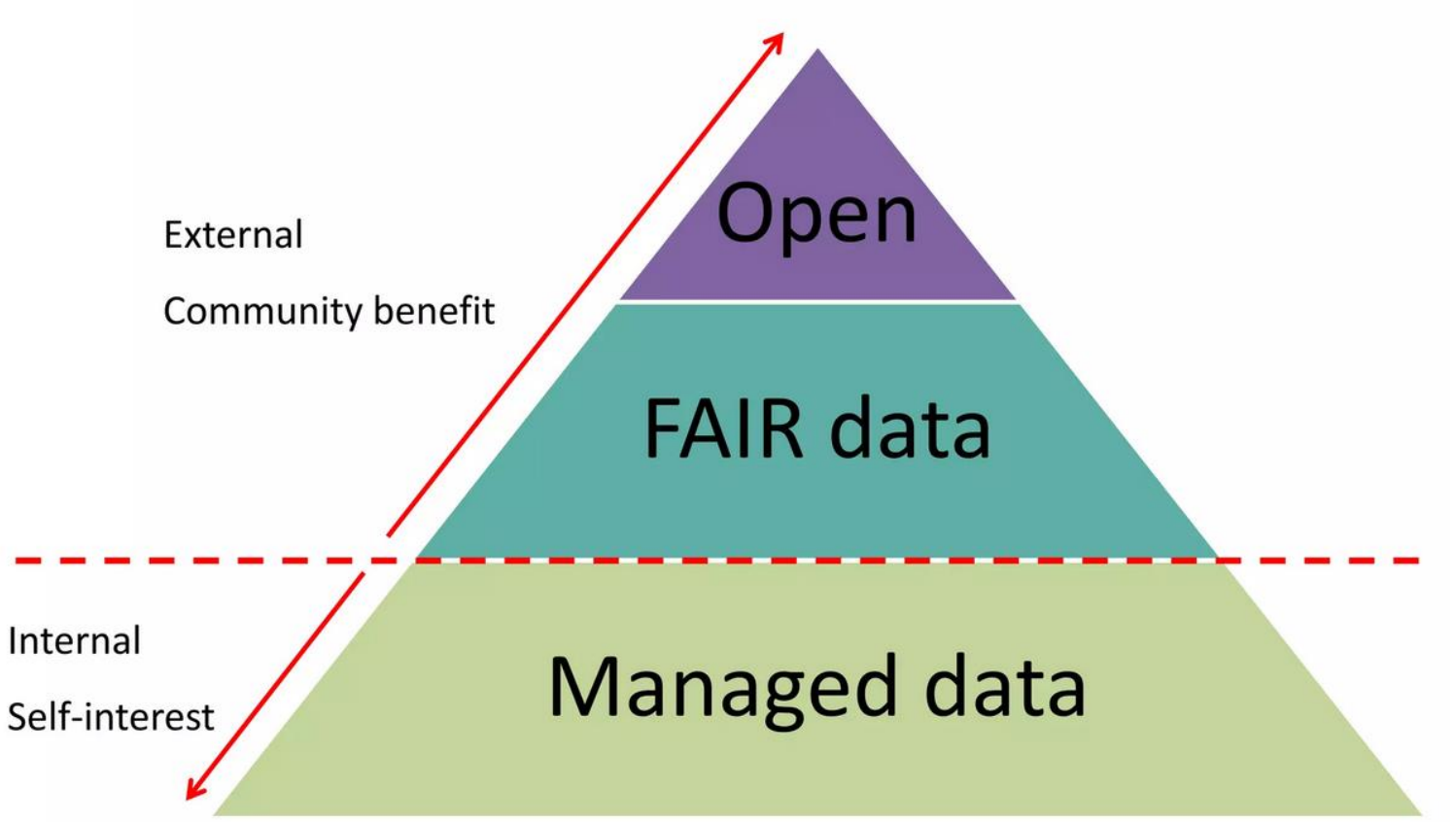
>  **Kaitlyn M. Werner, PhD** @kaitlynmwerner · May 1
>
> I have been thinking a lot about socioeconomic status and self-control/self-regulation. I'm starting to plan an esm+diary study where I can start digging into this topic in more detail, but in the meantime I'm curious: what are the interesting papers you've read in this space?
>
> Show this thread

11:52 PM · May 9, 2022 · Twitter Web App

Qualitative research?

I was born in [redacted]. My parents were both born and raised in [redacted]. My father, [redacted], was [redacted] and my mother, [redacted], was [redacted]. They both lived in [redacted], on either side of [redacted], and there was no chance that they would meet each other.

As **open** as possible, as **closed** as necessary



Source: https://www.slideshare.net/sjDCC/open-fair-data-and-rdm

Collective benefits

More robust research

Verification of results

New collaborations
(across disciplines and borders)

New uses of data

No duplications

Use of data in teaching

Individual benefits

Increased visibility

More data reuse

New collaborations

Increased citations

Collective benefits

More robust research

Verification of results

New collaborations
(across disciplines and borders)

New uses of data

No duplications

Use of data in teaching

Requirements

Funders and publishers

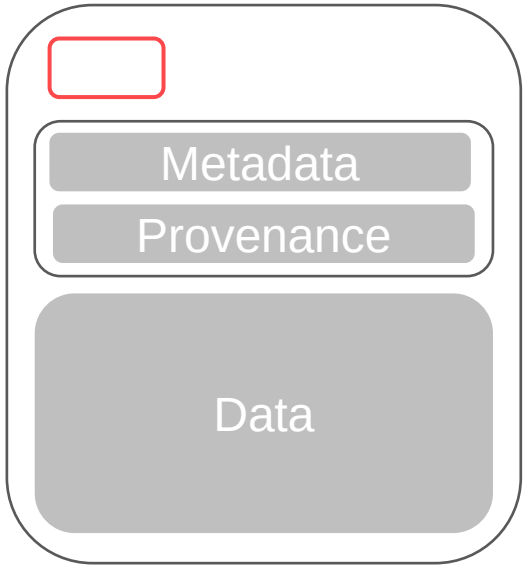
Institutions

New assessment systems

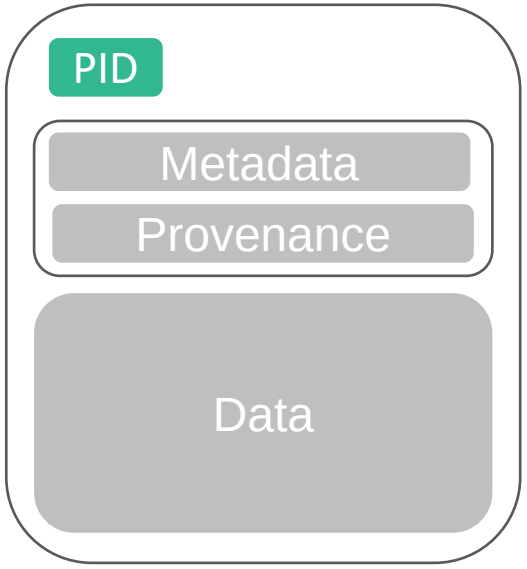
Findable Accessible Interoperable Reusable

Levels of FAIR

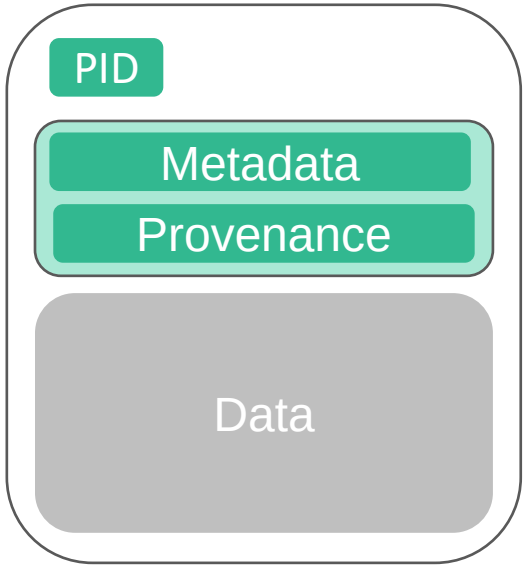
Totally UNFAIR



Findable
Usable for humans



FAIR metadata



Adapted from:
https://www.force11.org/

https://twitter.com/mads100tist/status/1366103674989277185

Data availability statements don't work

Current Biology 24, 94–97, January 6, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.11.014

Report

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines,[1,2,*] Arianne Y.K. Albert,[3] Rose L. Andrew,[1] Florence Débarre,[1,4] Dan G. Bock,[1] Michelle T. Franklin,[1,5] Kimberly J. Gilbert,[1] Jean-Sébastien Moore,[1,6] Sébastien Renaut,[1] and Diana J. Rennison[1]

[1]Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada
[2]Molecular Ecology Editorial Office, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada
[3]Women's Health Research Institute, 4500 Oak Street, Vancouver, BC V6H 3N1, Canada
[4]Centre for Ecology & Conservation Biosciences, University of Exeter, Cornwall Campus, Tremough, Penryn TR10 9EZ, UK
[5]Institute for Sustainable Horticulture, Kwantlen Polytechnic University, 12666 72nd Avenue, Surrey, BC V3W 2M8, Canada
[6]Department of Biology, Université Laval, 1030 Avenue de la Médecine, Laval, QC G1V 0A6, Canada

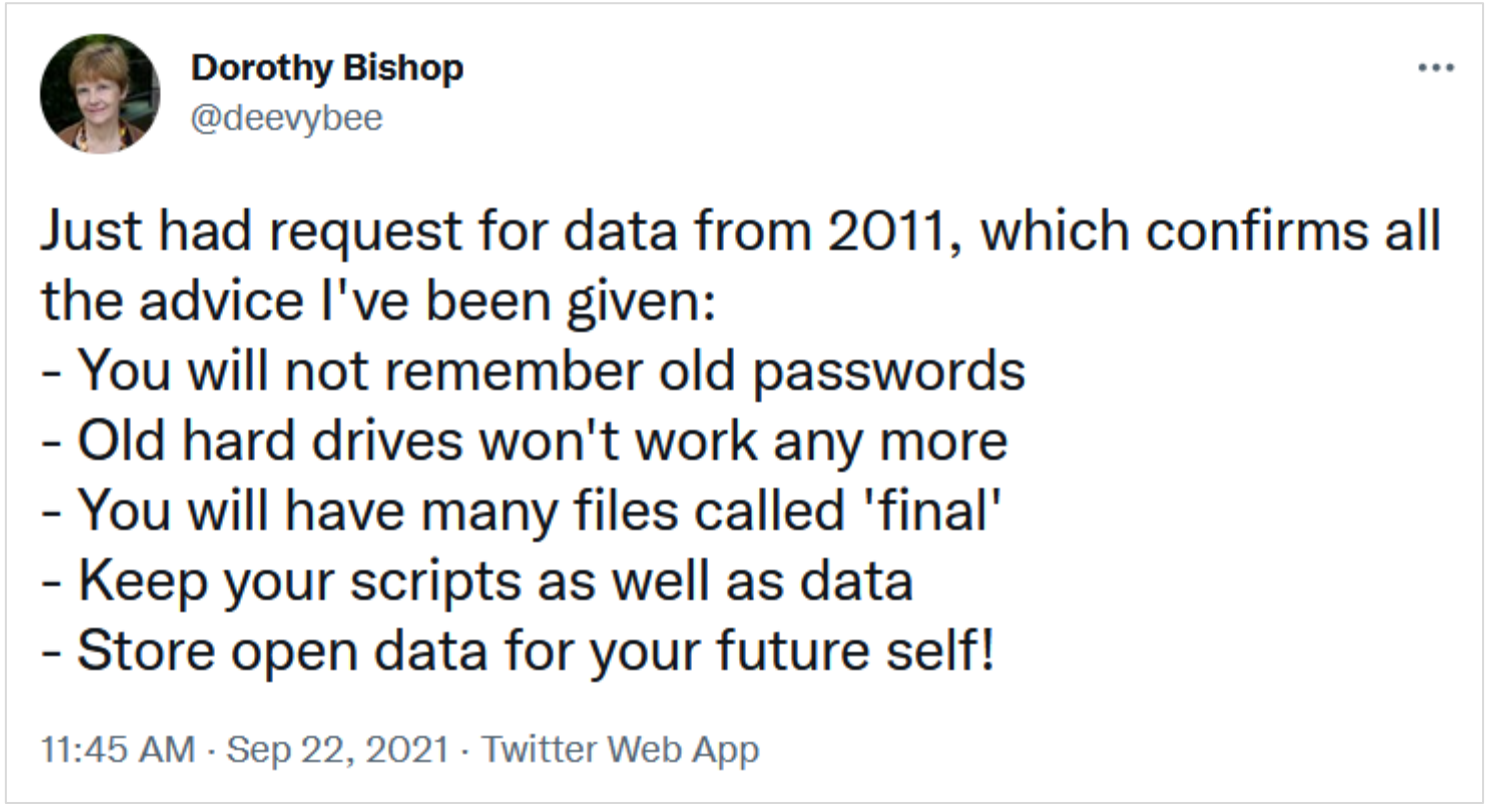
Summary

Policies ensuring that research data are available on public archives are increasingly being implemented at the government [1], funding agency [2–4], and journal [5, 6] level. These policies are predicated on the idea that authors are poor stewards of their data, particularly over the long term [7], and indeed many studies have found that authors are often unable or unwilling to share their data [8–11]. However, there are no systematic estimates of how the availability of research data changes with time since publication. We therefore requested data sets from a relatively homogenous set of 516 articles published between 2 and 22 years ago, and found that availability of the data was strongly affected by sets (23%) were confirmed as extant. Table 1 provides a breakdown of the data by year.

We used logistic regression to formally investigate the relationships between the age of the paper and (1) the probability that at least one e-mail appeared to work (i.e., did not generate an error message), (2) the conditional probability of a response given that at least one e-mail appeared to work, (3) the conditional probability of getting a response that indicated the status of the data (data lost, data exist but unwilling to share, or data shared) given that a response was received, and, finally, (4) the conditional probability that the data were extant (either "shared" or "exists but unwilling to share") given that an informative response was received.

There was a negative relationship between the age of the paper and the probability of finding at least one apparently working e-mail either in the paper or by searching online (odds ratio [OR] = 0.93 [0.90–0.96, 95% confidence interval (CI)], $p < 0.00001$). The odds ratio suggests that for every year since publication, the odds of finding at least one apparently working e-mail decreased by 7% (Figure 1A). Since we searched for e-mails in both the paper and online, four factors contribute to the probability of finding a working e-mail: (1) the number of e-mails in the paper and (2) the chance that any of those worked and (3) the number of e-mails we could find by searching online and (4) the chance that any of those worked. The total number of e-mail addresses we found in the paper decreased with age (Poisson regression coefficient = −0.07, SE = 0.01, $p < 0.0001$) from an average of 1.17 in 2011 to 0.42 in 1991 (Figure 2A), and there was a slight positive effect of article age on the number of e-mails we found online (Poisson regression coefficient = 0.015, SE = 0.007, $p < 0.05$; Figure 2C). Moreover, the chance that an e-mail found in the

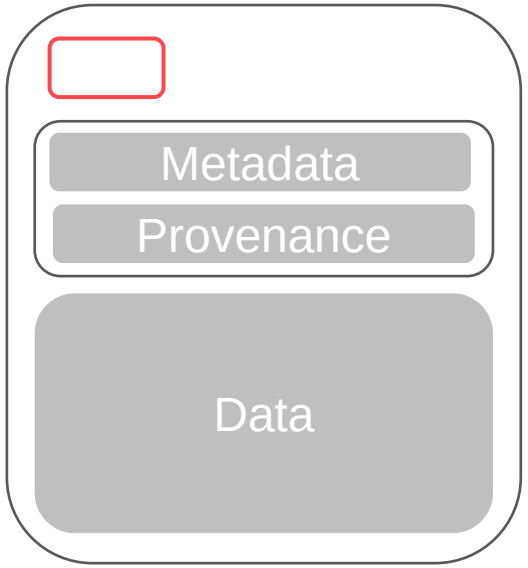
Data availability statements don’t work



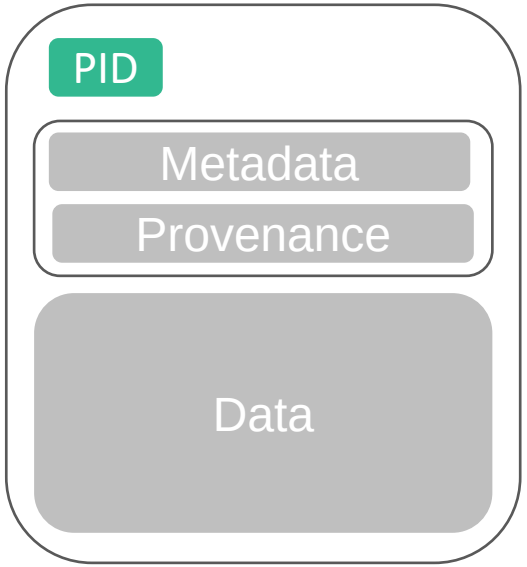
https://twitter.com/deevybee/status/1440613146788384768

Levels of FAIR

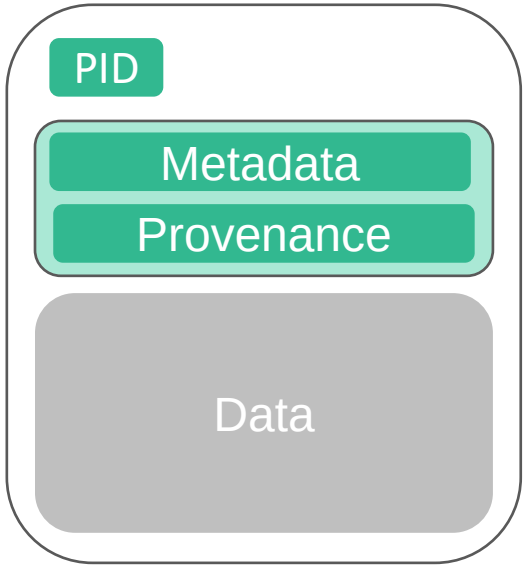
Totally UNFAIR



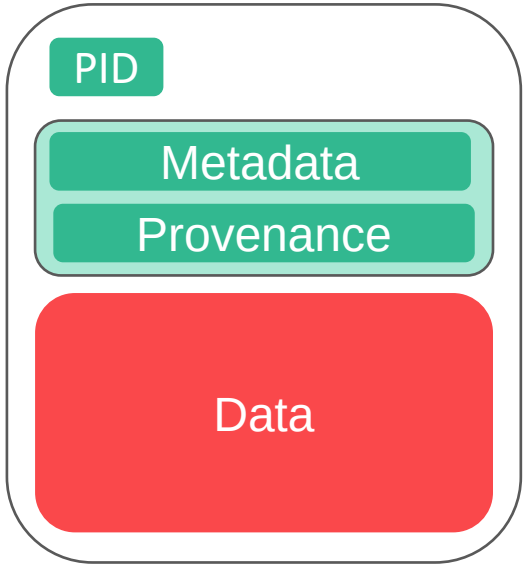
Findable
Usable for humans



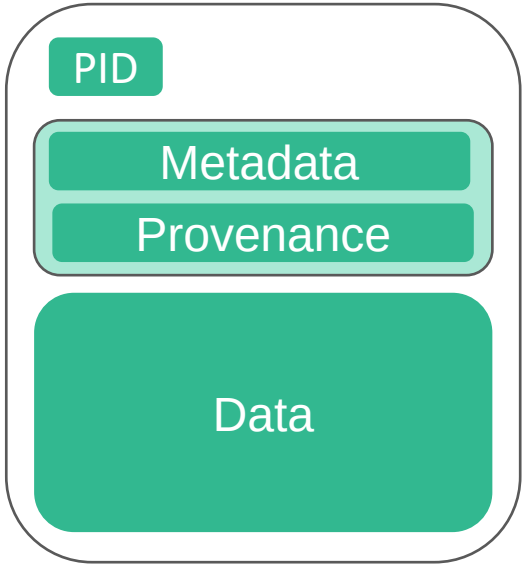
FAIR metadata



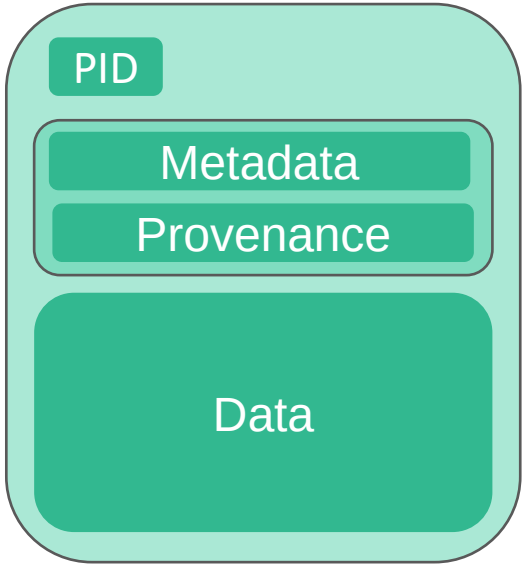
FAIR data
Restricted access



FAIR data
Open access



FAIR data
Open access and functionally linked



Adapted from:
https://www.force11.org

1) Presentation from **Centre for Cancer Cell Reprogramming (CanCell)**
   Associate professor **Ragnhild Eskeland**, University of Oslo

2) Presentation from **Department of Social Anthropology (SAI)**
   Professor **Marianne Elisabeth Lien**, University of Oslo

3) Presentation from **Nordic Centre of Excellence Quality in Nordic Teaching (QUINT)**
   Professor og centre leader **Kirsti Klette**, University of Oslo