Changing incentive structures to foster the actual sharing rate of open data

or:

Why everybody loves data sharing, but nobody does it.
Open Science
“the European Commission is now moving decisively from ‘Open access’ into the broader picture of ‘Open science’”

→ Open Data is default (with opt-out possibility)

• German Research Foundation (DFG): Publicly funded research data belongs to the public

• G7 science ministers: „recognize open science practices during evaluation of funding proposals and outcomes; reward open science activities in career advancement“

The Mertonian norms of science

Communality
The findings of science belong to everyone, they are not private property.

Organized skepticism
All ideas must be tested and are subject to rigorous, structured community scrutiny.

Disinterestedness
Scientists should be focused on finding the truth, not on their own success.

Universalism
The validity of a scientific claim does not depend on who is making it.

Counternorm: Secrecy
Hiding procedures, materials, and results

Counternorm: Organized Dogmatism:
Old findings are not challenged, no independent verification takes place.

The Mertonian norms of science

FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).


N = 3,247
MUTUAL TRUST RELATIONSHIP

Researchers

Society

Funders

Economy
Everybody else in academia

#osc2018
#oscibar
The Mertonian norms of science

FIG. 3. Norm versus Counternorm Scores: Percent with Norm > Counternorm (dotted), Norm = Counternorm (striped), Norm < Counternorm (solid).


N = 3,247
Eroding trust in science

IS THERE A REPRODUCIBILITY CRISIS?

1,576 researchers surveyed

3% No, there is no crisis
7% Don’t know
52% Yes, a significant crisis
38% Yes, a slight crisis

90%: Yes
„Sharing upon request“ as a policy is dead

- **100%** of authors in these studies signed to share the data upon request
- Actual sharing rate (Wicherts et al., 2006): **27%** (out of 141 requests)
- Vanpaemel et al. (2015): **38%** (out of 394 requests)
- Stodden et al. (2018): **44%** (out of 204 requests) provided some „artifacts“, 26% could be reproduced

- Bus factor / long-term availability?
- Providing selective access (e.g., not to critics)?
- Data set providers should not be in charge for access control ➔ either fully open, or independent stewards grant access based on prespecified rules
Why not sharing
# Rewarding quantity, not quality

## Actual (not desired) relevance in professorship hiring committees

<table>
<thead>
<tr>
<th></th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong> of peer-reviewed publications</td>
<td>1</td>
</tr>
<tr>
<td>Fit of research profile to the hiring department</td>
<td>2</td>
</tr>
<tr>
<td>Quality of research talk</td>
<td>3</td>
</tr>
<tr>
<td><strong>Number</strong> of publications</td>
<td>4</td>
</tr>
<tr>
<td><strong>Volume</strong> of acquired third-party funding</td>
<td>5</td>
</tr>
<tr>
<td><strong>Number</strong> of first authorships</td>
<td>6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

N = 1453 psychology researchers, 66% were members of a professorship hiring committee.

Early career researchers are stuck

What would be a good balance between Open Science and having a career in academia? […] Being open IMHO is a competitive disadvantage. Can you only afford open science when you are tenured?

Why should I share my hard-won data with my rivals that presumably compete with me for the next post-doc position?

My contract is limited to two years – although it would be nice to publish the data, I have no time to do it. I rather have to churn out another publication.

➡️ felt contradiction between „good research“/„open research“ and „having a career in science“
Quantity, not quality

<table>
<thead>
<tr>
<th>Actual (not desired) relevance in professorship hiring committees</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of peer-reviewed publications</td>
<td>1</td>
</tr>
<tr>
<td>Fit of research profile to the hiring department</td>
<td>2</td>
</tr>
<tr>
<td>Quality of research talk</td>
<td>3</td>
</tr>
<tr>
<td>Number of publications</td>
<td>4</td>
</tr>
<tr>
<td>Volume of acquired third-party funding</td>
<td>5</td>
</tr>
<tr>
<td>Number of first authorships</td>
<td>6</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>Quality rating of the three best publications</td>
<td>17</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

**Indicators of research transparency** | 41 (of 41)

N = 1453 psychology researchers, 66% were members of a professorship hiring committee.

It is the responsibility of senior researchers, funders, and policy makers to resolve this social dilemma for young researchers.
Quantity, not quality

Highest discrepancies between desired relevance and actual relevance

N = 1453 psychology researchers, 66% were members of a professorship hiring committee.

The typical researcher’s narrative about data sharing / open science
maybe slightly exaggerated (but maybe not)

• Nobody does it – why should I?
• A lot of work, which is not rewarded.
• RDM is BORING
• Strategic trade-off: More papers on CV, or documenting old stuff? In order to get tenure/more grant money, I’d rather optimize the former.
• Please: No bureaucratic over-regulation. Protect academic freedom!
Going forward: How to increase the actual sharing rate
How to achieve cultural change

1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is "good" scientific practice

4: Incentives
Reward openness

5: Policy
Require openness

Barriers

No proper recognition for sharing (27%)

Sharing data is not a common practice in my field (68%)

Preparing data is too time-consuming (55%)

There is no suitable repository to share my data (12%)
I never learned to share data online (54%)  

http://doi.org/10.1177/2515245917751886
How to achieve cultural change

1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is “good” scientific practice

4: Incentives
Reward openness

5: Policy
Require openness

[Image of re3data.org and EOSC]
1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is "good" scientific practice

4: Incentives
Reward openness

5: Policy
Require openness

The implementation of reliability requires robust infrastructure. Interfaces make it easy to perform the behaviors. Norms define what is acceptable and good practice in scientific research. Incentives reward openness, encouraging the sharing of data and results. Policies require openness, ensuring that all members of the community adhere to these practices.

Dublin Core Metadata Scheme

3.1 URIs in DC-DS-XML

The Dublin Core Metadata Scheme (DCMS) uses Uniform Resource Identifiers (URIs) [RFC3986] to refer to both metadata and resources. These URIs are encoded as XML attributes and can be abbreviated through the use of XML entities. The following example shows a URI as the value for the dc:identifier and dc:title attributes:

```xml
<dc:title>Home Page</dc:title>
```

In DC-DS-XML, URIs are encoded as attributes, used as XML attribute values, and in different XML attributes in detail. The purpose of this section is to make some general statements about DC-DS-XML URI attributes. The URI may be represented in full. The following example shows a URI as the value:

```xml
<dc:identifier encoding="UTF-8">http://purl.org/dc/terms/title</dc:identifier>
```
1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is “good” scientific practice

4: Incentives
Reward openness

5: Policy
Require openness

https://datawiz.leibniz-psychology.org ➞ Software solutions + supporting persons (data stewards) at the local level
We suggest that beginning January 1, 2017, reviewers make open practices a pre-condition for more comprehensive review. This is already in reviewers’ power; to drive the change, all that is needed is for reviewers to collectively agree that the time for change has come.
• More and more journals change from an opt-in to an opt-out (+public justification) policy

• Educate students: This is how science is done - these are the norms of good scientific practice and integrity.
Open Science Badges

5: Policy
Require openness

4: Incentives
Reward openness

3: Norms
Communities define and communicate what is "good" scientific practice

2: Interfaces
Workflows that make it easy to do the behaviors

1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

<table>
<thead>
<tr>
<th>Extra cost for journals</th>
<th>Extra cost for reviewers</th>
<th>Extra cost for researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very few (add badges to workflow)</td>
<td>Few (verify availability) to some (reproduce)</td>
<td>Some</td>
</tr>
</tbody>
</table>
Open Science Badges

1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is "good" scientific practice

4: Incentives
Reward openness

5: Policy
Require openness

---

### Extra cost for journals
- Very few (add badges to workflow)

### Extra cost for reviewers
- Few (verify availability) to some (reproduce)

### Extra cost for researchers
- Some

---

As of Oct 2015, 38% of all PsychScience papers had Open Data

https://osf.io/tvyxz/wiki/home/
Kidwell et al. (2016). http://doi.org/10.1371/journal.pbio.1002456
Funders: Add „Public data sets“ as a section to CV templates

Suggestion:
„Publication list must include a section with up to 5 of the most impactful public data sets that an applicant provides, with a one-sentence statement about each data set’s specific impact."

Extra cost for funders: None (add a few sentences to guidelines)
Extra cost for reviewers: None (take information into consideration)
Extra cost for researchers: ~5 min.

1.1 Publications list as part of the academic curriculum vitae:
- must be included for each applicant,
- need not directly relate to the proposed project,
- must include up to ten of the most important publications for each applicant,
Hiring committees: Make „open science“ a desirable or essential job characteristic

Since 2015: All professorship job descriptions use this requirement

See more such prof job ads at: https://osf.io/7jbnt/
**Hiring committees: Make „open science“ a desirable or essential job characteristic**

For staff roles involving **at least some research**, signatories (employers) self-certify to meet ONE of the levels below. Signatories may wish to apply different levels of commitment for different grades or type of appointment. **Typical categories could be (a) PhD students/research assistants, (b) Post-Doc, or (c) faculty (i.e., associate and full professors).**

<table>
<thead>
<tr>
<th>Extra cost for committees</th>
<th>None (add a paragraph to job description)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra cost for reviewers</td>
<td>None (take information into consideration)</td>
</tr>
<tr>
<td>Extra cost for applicants</td>
<td>a few minutes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 0</th>
<th>Level I</th>
<th>Level II</th>
<th>Level III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual or organisation makes no commitment to mention open research practices in published hiring policies or advertised research job descriptions.</td>
<td>Individual or organisation makes no commitment to mention open research practices in advertised research-related job descriptions but does include them as desirable characteristics in published hiring policy. All else being equal, candidates with greater open science track records may be preferred over candidates with no or lesser open science track record.</td>
<td>Individual or organisation commits to including proven track record of open research practices as desirable characteristics (but not necessarily as essential characteristics) in all advertised research-related job descriptions. All else being equal, candidates with greater open science track records are preferred over candidates with no or lesser open science track record.</td>
<td>Individual or organisation commits to including proven track record of open research practices in all advertised research-related job descriptions as essential characteristics. Only candidates with an open science track record are interviewed and/or appointed. All else being equal, candidates with greater open science track records are preferred over candidates with lesser open science track record.</td>
</tr>
</tbody>
</table>
Hiring committees:
Require an annotated CV with limited items (e.g., <= 10)

<table>
<thead>
<tr>
<th>Authors &amp; title</th>
<th>Year</th>
<th>Sample size per study</th>
<th>p-value per study</th>
<th>Open Science indicators</th>
<th>Data set</th>
<th>Applicants contribution</th>
</tr>
</thead>
</table>
| Doe, John & Smith, Peter | 2001 | n₁ = 21  
 n₂ = 30  
 n₃ = 19 | p₁ = .048  
 p₂ = .050  
 p₃ = .023 | ☐ Open Data  
 ☐ Open Material  
 ☑ Preregistered | ☑ Own data collection ➔ URL | • Analyzed data  
 • Wrote manuscript |
| Doe, John | 2016 | n₁ = 180  
 n₂ = 158 | p₁ = .012  
 p₂ = .001 | ☑ Open Data  
 ☑ Open Material  
 ☑ Preregistered | ☑ Own data collection ➔ URL  
 osf.io/as1cd  
 ☐ Archival data | • Designed study  
 • Wrote manuscript |

No journal; JIF is irrelevant or misleading

Paper-level citation metrics

Basic information for judging evidential value

Open science indicators: Judging replicability

Data: own collection or reuse?
“How likely are you to share your research data if . . .?”

1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is “good” scientific practice

4: Incentives
Reward openness

Who is targeted?
- Idealists
- Supporters
- Substantial middle looking for guidance
- Pragmatists

Pyramid based on a tweet storm by @BrianNosek
“How likely are you to share your research data if . . .?”

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>your research funder requires you to share</td>
<td>4%</td>
</tr>
<tr>
<td>the journal requires you to share</td>
<td>4%</td>
</tr>
<tr>
<td>your institution requires you to share</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>93%</td>
</tr>
<tr>
<td></td>
<td>94%</td>
</tr>
</tbody>
</table>
How to achieve cultural change

1: Implementation
Reliable infrastructure that makes it possible to do the behaviors

2: Interfaces
Workflows that make it easy to do the behaviors

3: Norms
Communities define and communicate what is „good“ scientific practice

4: Incentives
Reward openness

5: Policy
Require openness

„We expect our researchers to maximise the availability of research data, software and materials with as few restrictions as possible. As a minimum, the data underpinning research papers should be made available to other researchers at the time of publication. [...]"

Wellcome will also consider whether researchers have managed and shared their research outputs in line with our requirements, as a critical part of the end of grant reporting process.“

„expects and supports the timely release and sharing of final research data“

„erwartet der SNF, dass Daten [...] auf öffentlich zugänglichen, digitalen Datenbanken archiviert werden“

„It is recommended to make all research data [...] available for reuse, for example under Creative Commons licence“

Input control ➔ output control?
Funders: Require Transparency and Openness (TOP) statement in final reports

<table>
<thead>
<tr>
<th>Extra cost for funders</th>
<th>None (add a few sentences to guidelines)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra cost for reviewers</td>
<td>None (take information into consideration)</td>
</tr>
<tr>
<td>Extra cost for researchers</td>
<td>~5 min.</td>
</tr>
</tbody>
</table>

### Are the relevant data from the funded project accessible in an open repository?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Provide a persistent, unique identifier and any required instructions</td>
</tr>
<tr>
<td>No</td>
<td>Provide justification (short free text)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>Provide explanation (short free text)</td>
</tr>
</tbody>
</table>

### Have you cited any previously generated data used in this project?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Not applicable</td>
<td></td>
</tr>
</tbody>
</table>

1. Disclose ➔
2. Require ➔
3. Verify

Action List: „Bridging the last mile“

• **Universities**: Educate and practice the values and principles of good scientific practice.

• **Universities**: Provide supporting infrastructure, such as data stewards.

• **Universities**: Add research transparency as desirable or essential job characteristic for post-doc and prof positions

• **Infrastructure**: Provide *user-friendly* tools

• **Journals**: Make open data the default; authors can opt-out with a *public* justification

• **Funders**: Appreciate openness in grant proposal (both on project level and applicant level)

• **Funders**: Require transparency and openness statement in final reports; use openness track record for future decisions
Fast adoption vs. High (FAIR) quality?

• Low hurdles, one small step at a time

• Reward small steps
  *Sharing something - even badly documented data - is better than sharing nothing.*

• Learning by doing
  *With increasing practice, hopefully the quality gets better, too.*

• But: (Initially) Low quality
  *Barely reusable data sets; trying to reproduce a result is a pain in the ass or impossible; data reuse very limited.*

• Risk of „open-washing“
  *Pretending openness without actual value.*

• High hurdles
  *Mainly enthusiasts/computer scientists will able and motivated use it*

• Reward big steps
  *Curated repositories with input quality control.*

• Instant high quality
  *The data sets which are open are instantly FAIR.*
Thanks

https://commons.wikimedia.org/wiki/File:Soap_Bubble_-_foliage_background_-_iridescent_colours_-_Traquair_040801.jpg