Data Sharing in Academia

Marcel Hebing (DIW), presenting joint work with Benedikt Fecher (HIIG), Sascha Friesike (HIIG), and Stephanie Linek (ZBW) at IASSIST 2015 in Minneapolis
Agenda

1 Framework for data sharing in academia

2 Study design

3 First results: it's a reputation economy
1 Framework for Data Sharing in Academia

Design of the preparatory study

1. Systematic review of 97 scholarly papers concerning data sharing
   → qualitative content analysis and meta analysis
   → identification of factors + structure
2. Preliminary category system
3. Quantitative study of secondary data users (SOEP User Survey)
   → 603 participants
   → opened and closed questions regarding data sharing
   → validation of preliminary system
4. Framework for academic data sharing
Data donor 125 (101/24)
- Sociodemographic factors
- Degree of control
- Resources needed
- Returns

Norms 85 (74/11)
- Ethical norms
- Legal norms

Data infrastructure 76 (67/9)
- Architecture
- Usability
- Management software

Research org. 50 (50/0)
- Data donor’s organization
- Funding agencies

Research community 117 (75/60)
- Data sharing culture
- Standards
- Scientific value
- Publications

Data recipients 88 (55/33)
- Adverse use
- Recipient’s organization

Non-academic recipient

Academic recipient

Data flow

Category name
Total num. of references (from review/survey)
Online resources

Article:

doi:10.1371/journal.pone.0118053

Data:

doi:10.5684/dsa-01
2 Study design
Questionnaire

Topics:

→ Working environment
→ Research practices and handling of data
→ Sharing data: enablers and barriers based on the framework
→ Sharing in general
→ Using secondary data
→ Social demographics including personality traits
Methodology

→ Time frame: September/October 2014
→ Quantitative survey of researchers
→ Most participants are from Germany
→ Sample size: approximately 1500
Sample statistics

**Status**
- Researchers without PhD: 12%
- Professors: 20%
- PhD students: 26%
- Researchers with PhD: 40%
- Students: 2%

**Disciplines**
- Natural sciences: 33%
- Human sciences: 12%
- Social sciences: 31%
- Humanities: 11%
- Engineering: 8%
- Agriculture: 5%
Online resources

Questionnaire (in English and German):

doi:10.5684/dsa-02

Working paper:

http://ratswd.de(dl)/RatSWD_WP_246.pdf

LSE blog:

http://blogs.lse.ac.uk/impactofsocialsciences/
3 First results: It's a reputation economy.

First results

→ Willingness to share unconditionally is high in the natural and social sciences, and humanities (22-31%)—and low in engineering and medicine (11-15%).
→ I know where and how to share data. While 60% in the natural sciences agree, only 42% in social sciences.
→ Male are much more likely to share data (almost twice as much)
Enablers

I would only share my data...

→ if I knew what the data were going to be used for. (3.27)
→ if sharing the data enabled me to get into contact with other researchers. (3.04)
→ if I had enough time beforehand, to publish on the basis of my data. (4.10)
→ if I knew who would be able to access the data. (3.18)
→ if my employer supported me actively. (3.59)
→ if I were cited in publications using my data. (4.21)
→ if I were given a co-authorship of articles using my data. (2.86)
→ if I received financial compensation for the effort. (2.19)
Enablers

I would only share my data...

→ if I knew what the data were going to be used for. (3.27)
→ if sharing the data enabled me to get into contact with other researchers. (3.04)
→ if I had enough time beforehand, to publish on the basis of my data. (4.10)
→ if I knew who would be able to access the data. (3.18)
→ if my employer supported me actively. (3.59)
→ if I were cited in publications using my data. (4.21)
→ if I were given a co-authorship of articles using my data. (2.86)
→ if I received financial compensation for the effort. (2.19)
Enablers

I would only share my data...

→ if I knew what the data were going to be used for. (3.27)
→ if sharing the data enabled me to get into contact with other researchers. (3.04)
→ if I had enough time beforehand, to publish on the basis of my data. (4.10)
→ if I knew who would be able to access the data. (3.18)
→ if my employer supported me actively. (3.59)
→ if I were cited in publications using my data. (4.21)
→ if I were given a co-authorship of articles using my data. (2.86)
→ if I received financial compensation for the effort. (2.19)
Barriers

I would not share my data…

→ if other researchers could use my data to publish before me. (4.25)
→ if others could criticize or falsify my work. (1.99)
→ if the data could be misinterpreted. (3.16)
→ if the data collection required considerable effort. (2.57)
→ if a major effort were required to share the data. (3.55)
Barriers

I would not share my data…

→ if other researchers could use my data to publish before me. (4.25)
→ if others could criticize or falsify my work. (1.99)
→ if the data could be misinterpreted. (3.16)
→ if the data collection required considerable effort. (2.57)
→ if a major effort were required to share the data. (3.55)
Barriers

I would not share my data…

→ if other researchers could use my data to publish before me. (4.25)
→ if others could criticize or falsify my work. (1.99)
→ if the data could be misinterpreted. (3.16)
→ if the data collection required considerable effort. (2.57)
→ **if a major effort were required to share the data.** (3.55)
Thank you!

Marcel Hebing (DIW Berlin)
marcel@ddionrails.org
www.data-sharing.org