**Research Data Management**

## A European Perspective

Edited by

Filip Kruse and Jesper Boserup Thestrup

TECHNISCHE INFORflt.ATIO !S81SL'OT:-lEK (TIB}



**DE GRUYTER**

## SAUR

subito e.V. licensed customer copy supplied for Wageningen University & Research - Library (SLI03X00597E)

**Veerle Van den Eynden**

**3 What Motivates Researchers to Manage and Share Research Data**

**Abstract:** Researchers are motivated to implement good research data management practices for a variety of reasons. Often, it is because they need to create research data that have wide and long-term usability, and that provide evidence to support published research fmdings. Collaboration in research is therefore a strong motivator: if collaborators need to be able to understand and analyse one's data, the data need to be managed accordingly. Also data sharing and publishing open data strongly encourage good data management practices. In this case motivations are then driven by the incentive to share research data. Four recent studies about such incentives for researchers to share their data, indicate that both individual incentives and institutional or contextual factors play a role in motivating researchers to share their data. Direct research benefits, career benefits and rewards are strong motivators. Also the data sharing norms within a research group, community or discipline are influential. Data sharing policies, infrastructure and data services can provide a framework to support re­ searchers with data management and thereby contribute towards good research data management practices.

1. **lntroduction**

What motivates researchers to manage research data well? And what motivates them to share those data with other researchers or to make them available for public use? Working at a digital data archive where social science data are curated and disseminated, we know that these two aspects are aften closely linked. Making research data available for future reuse (sharing them) certainly requires a good level of data management. When data are deposited with us by researchers, we see a mix of how well (or not) these have been managed before being deposited. So what makes one person implement excellent data management, and another not?

Data management refers to all aspects of handling, housing, maintaining

and preserving data, that together ensure that data are of high quality, well organised, clearly documented, preserved and accessible and their validity con­ trolled at all times. This then facilitates efficient and high quality research based

DOi 10.1515/9783110365634-004

subito e.V. licensed customer copy supplied for Wageningen University & Research - Library (SLI03X00597E)

on solid data. One would think that the desire for high quality research alone would motivate any researcher to implement good data management practices. That is not necessarily the case. Data management also requires time and effort, which may compete with other research activities such as publishing. So at times there may be a trade-off between data management activities and other research activities. A researcher also needs the skills and tools to implement good data practices (Corti and Van den Eynden 2015).

Besides this immediate individual research need, there are other - more institutional - drivers for good data management. The expectations around data publication and data sharing are probably the most important drivers for good data management. Data sharing is increasingly encouraged or required by research funders and journal publishers, but also from within the research community itself. Research funders want to maximise the scientific outputs and benefits to society from their investments, and making sure that data can be reused for future research plays an important role.

Concerns over potential fraud in science and the lack of replicability of published findings, bas pushed for more transparency about data on which claims are based. Journals therefore increasingly request data and analytical code that underpin published findings to be published as supplemental files or in a data repository. The lack of reproducibility is also a major concern to the

researchers themselves. Baker (2016) found that of 1576 researchers, 70% had tried and failed to reproduce other researchers' experiments and half had failed to reproduce their own research. Selective reporting, pressure to publish and poor statistics or analysis count as major factors. In line with those increasing requirements for data sharing and transparency, research institutions increas­ ingly implement data management policies that outline the responsibilities for researchers and the institution itself (Horton 2016).

With data sharing playing such prominent role in driving the need for good data management, various recent studies have investigated what motivates researchers to share data, and which incentives can encourage more sharing. The findings of four studies are described here as evidence for what motivates researchers to manage and share their data well. Three studies gathered quanti­ tative information via surveys and one used a qualitative interview and case study approach. The quantitative studies each started from known incentives and motivations reported in literature, to either rank them by importance or identify which factors are significant. The qualitative study elicited incentives and motivations from the interviewees, without preconceived knowledge.

# Evidence from research on incentives and motivations for data sharing

#### Qualitative cross-European study on incentives to share research data based on five case studies

Van den Eynden and Bishop (2014) investigated researchers' incentives for data sharing based on semi-structured in-depth interviews with 22 researchers of five research teams known to have a well-established data sharing culture. The five case studies span various academie disciplines: arts and humanities, social sciences, biomedicine, chemistry and biology, and were based in Denmark, Finland, Germany, the Netherlands and the United Kingdom. Cases were selected explicitly to explore known factors that influence data sharing, such as project scale and discipline.

Interviews covered a standard list of questions and focussed on the research done, the existing data sharing practices, motivations for researchers to share their data and incentives that would motivate more researchers to share data. The discussions about motivations did not start from any preconceived ideas about motivations. Interviewees were free to discuss anything they felt was influencing them in their decision to share their research data. Interviews were carried out by the investigators or by local interviewers and were audio-recorded, transcribed and translated into English. All interviews were summarised using a standard template to provide comparability of topics, and published (Van den Eynden and Bishop 2015). Motivations and incentives were coded and structured starting from the comparable summaries, through synthesis and comparison of the topics discussed across all interviews. This was followed by comparison and discussion of emerging themes and findings.

This research found that motivations for researchers to share research data could be grouped into four main categories: direct research benefits, direct career benefits, norms and external drivers.

In these case studies, researchers gain direct research benefi.ts from sharing

their data when data are shared for collaborative analysis of complex data, when research consists of the creation of public data resources, when research depends on data mining, when data are submitted to a joumal as evidence for published findings in a paper, and when data are shared for methods learning and development.

Researchers gain direct career benefits when data sharing gives visibility to

their work and leads to reciprocal data exchanges. An invitation to share data reassures researchers that their data are recognised as valuable by others.

The influence of norms was clear from the fact that researchers exposed to a culture of data sharing within their research circle or discipline, are very likely to adopt the same culture and consider data sharing as integral part of standard research practice.

External drivers shown to motivate data sharing are funder and publisher data sharing policies, data sharing infrastructures and services that support researchers with data management, and direct funding for data sharing projects. Evidence for each of these categories of motivations was provided by inter­ views from multiple case studies, showing that motivations and related data sharing behaviour cross the disciplinary boundaries. Also comparing multiple interviews within cases showed significant variation in data sharing attitudes and practices, making it clear that broad discipline and research group alone do not determine all aspects of data sharing. Incentives to share were also found

to vary across a researcher's career trajectory.

A strong emerging theme to incentivise more data sharing in future was the view of researchers that creating alevel playing field for all researchers to share data and change the collective attitude towards sharing is important. This would mean that researchers spending time and resources on sharing data (often re­ quiring preparation and documentation) would not be disadvantaged compared to those not doing so. Policies, agreements and data sharing training can help establish that level playing field so data sharing would become part of standard research practices.

Detailed evidence can be found in the published report, interviews and summaries (Van den Eynden and Bishop 2014, Van den Eynden and Bishop 2015).

### Open research survey with UK and international researchers

Recently, a UK-based project investigating researcher's attitudes and behaviours towards open research, investigated data and code sharing practices and be­ haviour amongst biomedical, humanities and social science researchers funded by two major UK research funders (Van den Eynden et al. 2016). Evidence was gathered by inviting 3208 researchers to participate in an online survey. The survey instrument had been designed based on extensive literature review and expert input, in order to test existing perceptions and knowledge about data and code sharing practices, behaviour, harriers and motivations in a quantifiable manner.

Evidence provided by 842 respondents shows open research practices are increasing anc;i the benefits outweigh the harriers for most researchers. Known

**0% 20% 40% 60% 100%**

**My funder requires me to share my data (N=404) Journal expects data to be accessible (N=405)**

**My research community expects data sharing (N=405) lt is good research practica to share data (N=410)**

**lt enables collaboration and contribution by others (N=405) lt has public health benefits, e.g. disease outbreaks (N=390)**



**Ability to respond rapidly to public health emec-gencies (N= 85)**



**Ethica! obligatîon to participants to maxfmize benefits tor society (N=394)**

**Contributes to academie credentials (N=401) Enables validation and/or replication of my research (N=404) lmproved visibility tor my research (N=401)**

1 **can get credit and more citations by sharing data (N=394)**

**Not at all important** • **Slightly important** • **Moderately important** • **Very important** ■**Extremely important**

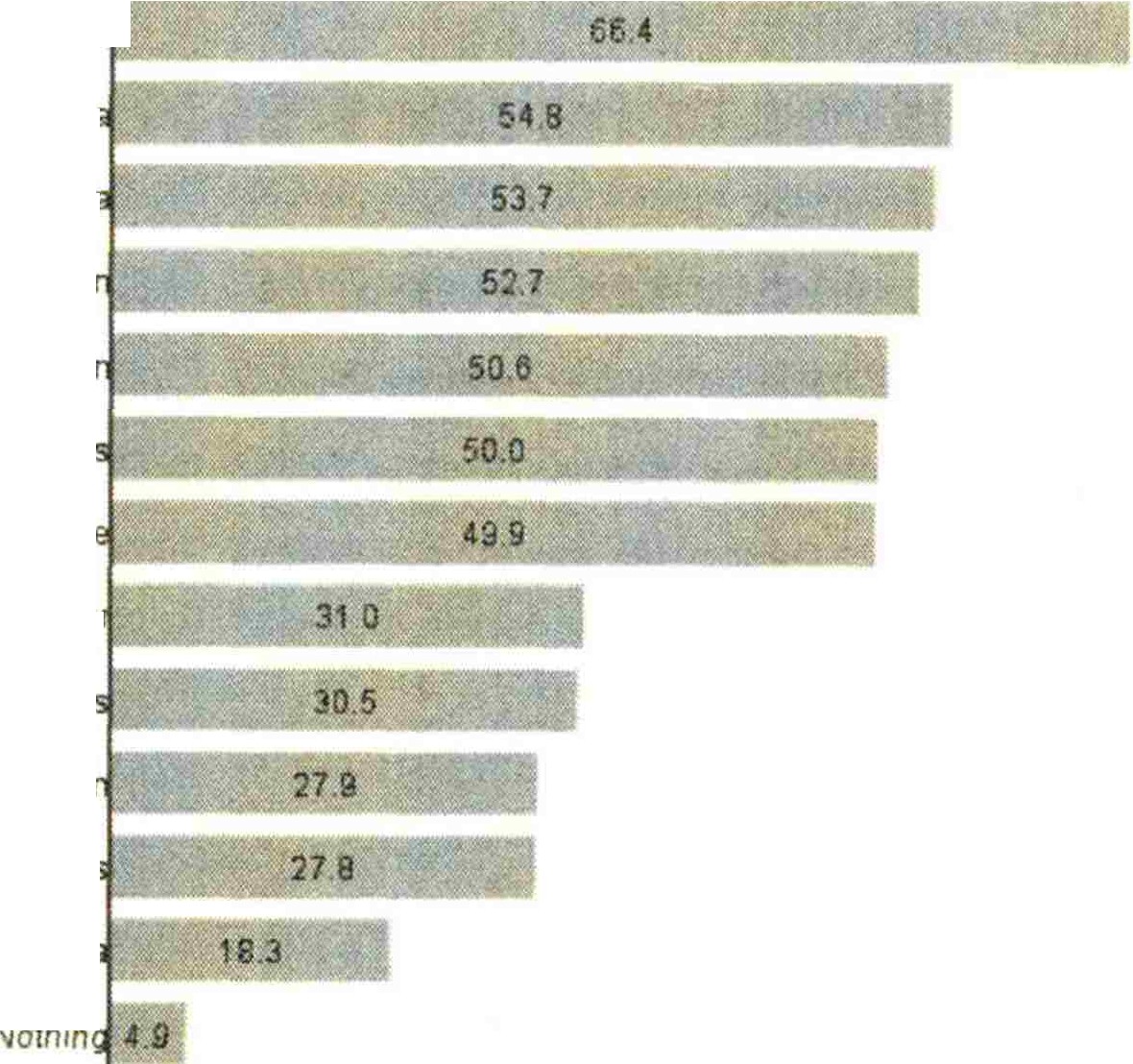
**Figure 3.1:** Reasons why researchers make research data available, scored on a 5-scale Likert scale by respondents from a list of factors previously reported in literature.

reasons and motivations for data sharing (from literature) were scored and ranked by respondents (Figure 3.1 and 3.2), but they could also add further reasons and motivations.

The principal reasons for respondents to share data are the fact that sharing data is seen to be good research practice, the requirement of a funder for data to be shared, to enable validation and replication of research, and because it enables collaboration and contribution by others (Figure 3.1). These reasons show significant variations according to career stage and discipline. For early career researchers, public health benefits, the ability to respond to health emer­ gencies, the ethica! obligation towards research participants and gaining citations and credit from data sharing are significantly more important reasons for sharing data. For humanities and social science researchers, funder requirements are a more important reason to share data.

Main factors that would motivate respondents to make more data available

in future are additional funding to cover the cost of sharing, knowing how re-users use their data, assistance from institutional or funder staff to prepare data for sharing and enhanced academie reputation (Figure 3.2). Humanities and social science researchers are more likely to be motivated to make more data available by additional funding, by case studies showcasing data, by the ability to control access to data and by assistance to prepare data for sharing. They are less likely to be motivated by enhanced academie reputation, by data

Extra funding to cover the cost

•

**Q)**

:ëi

-

**tQ**

tÖ

**"0**

..c

Knowtng how other people use my dat Asslstance rrom inst1tutionattfunder statrto prepare dat

Enhanced academie reputallo Evidence of data citat10

**m** Il is tooked on more ravourably in fund1ng and promotion decislon

Q)

**1***e***11**

Co-authorshlp on papers resultlng *trom* reus Data depos1t leads to pubhcatlon of a data pape

**til**

C

- iii

.i!:

ö

**:E**

Data access metric Flnanclal Incentive *trom* my inst1tutio

Ab1t1ty to limit data access to spec1fic purposes or individu.il

Case study that showcases my dat

0 20 40 60

**Percentage of respondent,;**

**Figure 3.2:** Factors that would motivate respondents to make available more data, selected by respondents from a list of factors previously reported in literature (N = 788).

access metrics, by data deposit resulting in publication of a data paper and by data sharing being looked upon more favourably in funding and promotion decisions. Such disciplinary different can be explained by the research methods and types of data: humanities and social science researchers are more likely to generate qualitative and textual data which may be more problematic to share when they contain personal and confidential information or less intuitive to share as reuse may seem less obvious.

Researchers who currently do not share their data are more likely to be motivated by the ability to control access to data and are more likely to state that nothing could motivate them to share data in future. Researchers who already make their data available are more likely to be motivated to share more data by financial incentives from their institution, by data access metrics and by case studies showcasing data.

Interestingly, very few respondents report having experienced any bad experiences from maldng data available, such as being scooped to publication, incorrect or inappropriate reuse of their data, reuse resulting in criticism, the time required to support re-users of data or receiving no acknowledgement from reuse.

#### Survey of STEM and social scientists in the US

Two similar USA-based studies with 1153 STEM scholars and 361 social science scholars investigated individual and institutional factors that influence data sharing across disciplines using the theory of planned behaviour and institu­ tional theory to predict data sharing behaviour (Youngseek and Stanton 2012; Youngseek and Adler 2015). The theory of planned behaviour provides insight into how an individual's attitudes, subjective norms and perceived behavioural control influence his/her behaviour. The individual factors believed to influence data sharing behaviour that were tested in these studies are: perceived career benefit, perceived career risk, perceived effort, attitude towards data sharing and scholarly altruism. Institutional theory believes that the institutional envi­ ronment produces a structured field of social expectations and norms, using

**Table 3.1:** Overview of individual and institutional factors that motivate researchers to share data, as reported in four studies

**Study lndividual factors lnstitutional factors**

Van den Eynden and Bishop (2014)

Van den Eynden et al. (2016): all disciplines

Van den Eynden et al. (2016): humanities and social science

Van den Eynden et al. (2016): early career researchers

Youngseek and Stanton (2012): STEM researchers

Youngseek and AdIer (2015): social scientists

Sayogo and Pardo (2013)

Direct research benefits Career benefits

Research benefits Knowledge of reuse Career benefits: enhanced academie reputation

Case studies showcasing data

Impact: public health benefits, respond to health emergencies Ethica[ obligation to research participants

Reward: citations and credit

Career benefits Scholarly altruism

Career benefits

Attitude to data sharing

Data management skills Rewards: citation, acknowledgement

Norms of research circle and/or discipline

Funder and journal poticies Data infrastructure and support services

Funding for data sharing

Norms: good research practice Funding for data management and sharing

Assistance for data management and sharing

Funding for data management and sharing

Assistance for data management and sharing Funder requirements

Norms

Journal requirements Norms

lnstitutional data management support

Legal/policy framework to

guarantee good reuse and acknowledgement

(dis)incentives to shape a person's beliefs, behaviour and practices. The institu­ tional factors believed to influence data sharing behaviour that were tested in these studies are: availability of data repositories, normative pressure, regulative pressure from journal publishers and regulative pressure from funding agencies. Evidence was gathered via two similar online surveys, sampling resp 16,165 and 2,285 random respondents from a database of USA academies.

These studies found that data sharing behaviour is mainly driven by personal motivations and perceived normative pressure, whilst pressure from funders and the availability of data repositories were not found to be significant motivators.

Amongst personal motivations for STEM scientists, perceived career benefits and scholarly altruism to share data were found to have a significant positive effect on data sharing, whilst perceived effort bas a significant negative effect. For social scientists, perceived career benefit and attitude towards data sharing have a positive effect, whilst perceived career risk and perceived effort have a negative effect on data sharing.

At an institutional level, normative pressure at discipline level bas a positive effect for both groups of scientists, and regulative pressure by journals bas a positive effect for STEM researchers.

##### Cross-disciplinary international survey on motivations for publishing research data

Sayogo and Pardo (2013) investigated researchers' motivations for publishing research data, via an online survey with 555 researchers from multiple dis­ ciplines. They assessed the willingness to publish datasets as open data against seven variables: organisational support, data management skills, data reuse acknowledgement, legal regulations and conditions of data reuse, concern for misinterpretation of data, economie motives and funder requirements.

Significant determinants for the likelihood that researchers will publish their research data were found to be: data management skills and institutional support for data management, and the acknowledgement of data reuse with a legal and policy framework that can guarantee proper reuse and attribution.

# Conclusion

These four studies carried out with different groups of researchers, in different academie and geographical settings, and using different methodologies, show that there is remarkable universality in what motivates researchers to share their

data so they can be used by other researchers and in future research. At the same time, they also point to some differences by research discipline and career stage. It is clear that certain incentives for data sharing operate at an individual level, whilst other factors are provided by an institutional context or by external influences such as funder and journal policies or the normative culture of the research discipline or setting. Table 3.1 gives an overview of the significant moti­ vations and incentives for data sharing identified by the four studies, organised into individual factors and institutional factors (which include all external factors imposed by external stakeholders such as joumal publishers, funding agencies, learned societies, and research institutions).

The three quantitative studies each started from known incentives and moti­ vations for data sharing, which were then either scored and ranked or tested for significant influences on data sharing. The results are therefore not fully comparable, as each study started from a different base list of factors. The qual­ itative study, however, observed and elicited incentives and motivations freely, and later grouped them into categories. This then provides the wider picture of which motivations may matter.

Amongst individual factors, direct career benefits and rewards such as citation and acknowledgement are important motivations for data sharing. Also direct research benefits play an important role.

Amongst institutional factors, the studies show that norms on data sharing that exist within a research discipline or research circle play a very important role in motivating researchers to share data (except Sayogo and Pardo (2013) who did not consider this factor). Support with data management and sharing, via funding, hands-on assistance and training to increase data management skills is shown to be influential too. Some studies indicate that funder and journal requirements for data publishing have an influence, whilst according to other studies this is not the case.

This knowledge about what motivates researchers to share their data can

be used by various stakeholders to develop strategies for promoting more publishing and sharing of research data. Developing rewards systems and careers benefits seems pivotal. The strong influence of data sharing norms implies a key role for early training on research data management a d sharing, for example as an integral part of research methods training. This also then influences researchers' data management skills and support, which in turn will influence da.ta sharing.

This brings us back to what motivates researchers to implement good data management practices. Increased expectations by various stakeholders for researchers to share and publish their data, requires good data management. We also see that data management skills and data management support (via funding or assistance) motivates researchers to share their data. The two are

clearly intricately linked. The same factors such as norms, research bene.fits, career bene.fits and the framework of policies, infrastructure and support would motivate researchers to implement good data management practices.

# References

Baker, Monya. 2016. "Is there a reproducibility crisis?" *Nature* 533: 452.

Corti, Louise and Veerle Van den Eynden. 2015. "Learning to manage and share data: jump­ starting the research methods curriculum," *International Journal of Socia/ Research Meth­ odology* 18(5): 545-559, accessed 01252017, doi: 10.1080/13645579.2015.1062627.

Horton, Laurence. 2016. *UK Higher Education lnstitution Research Data Management Policies, 2009-2016* (Data collection), Colchester: UK Data Archive, accessed 01252017, doi:10.5255/ UKDA-SN-851566.

Sayogo, Djoko S. and Theresa A. Pardo. 2013. "Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data," *Government Information Quarterly* 30(1): 19-31, accessed 01252017, doi:10.1016/j.giq.2012.06.011.

Van den Eynden, Veerle and Libby Bishop. 2014. *Sowing the seed: Incentives and Motivations for Sharing Research Data, a researcher's perspective.* Knowledge Exchange, accessed 01252017, http:// repository.j isc.ac.uk/5662/1/ KE\_repo rt-incentives-fo r-shari ng-researehdata.pdf.

Van den Eynden, Veerle and Libby Bishop. 2015. *Interviews of incentives for sharing research data.* (Data collection), Colchester: UK Data Archive, accessed 05312017, doi: 10.5255/ UKDA-SN-851540.

Van den Eynden, Veerle, Gareth Knight, Anca Vlad, Barry Radler, Carol Tenopir, David Leon, Frank Manista, Jim Whitworth, and Louise Corti. 2016. *Towards Open Research: practices, experiences, barriers and opportunities.* London: Wellcome Trust, accessed 01252017, doi: 10.6084/m9.figshare.4055448.v1.

Youngseek, Kim and Jeffrey M. Stanton. 2012. "lnstitutional and lndividual lnfluences on Scientists' Data Sharing Practices," *Journal of Computational Science Education,* 3(1): 47-56.

Youngseek, Kim and Melissa Adler. 2015. "Social scientists' data sharing behaviors: lnvestigat­ ing the roles of individual motivations, institutional pressures, and data repositories," *International Journal of Information Management* 35(4): 408-418, accessed 25012017, doi:10.1016 /j. iji nfomgt. 2015.04.007.