Reputation instead of obligation: forging new policies to motivate academic data sharing.

6/2/2015

Despite strong support from funding agencies and policy makers academic data sharing sees hardly any adoption among researchers. Current policies that try to foster academic data sharing fail, as they try to either motivate researchers to share for the common good or force researchers to publish their data. Instead, **Dr Sascha Friesike**, **Benedikt Fecher**, **Marcel Hebing***, and* **Stephanie Linek** *argue that in order to tap into the vast potential that is attributed to academic data sharing we need to forge new policies that follow the guiding principle reputation instead of obligation.*

In 1996, leaders of the scientific community met in Bermuda and agreed on a set of rules and standards for the publication of human genome data. What became known as the Bermuda Principles can be considered a milestone for the decoding of our DNA. These principles have been widely acknowledged for their contribution towards an understanding of disease causation and the interplay between the sequence of the human genome. The principles shaped the practice of an entire research field as it established a culture of data sharing. Ever since, the Bermuda Principles are used to showcase how the publication of data can enable scientific progress.

Considering this vast potential, it comes as no surprise that open research data finds prominent support from policy makers, funding agencies, and researchers themselves. However, recent studies show that it is hardly ever practised. We argue that the academic system is a reputation economy in which researchers are best motivated to perform activities if those pay in the form of reputation. Therefore, the hesitant adoption of data sharing practices can mainly be explained by the absence of formal recognition. And we should change this.

Useful but Hardly Practiced

The research landscape today is characterized by a collaboration imperative. Research questions are getting increasingly complex, and a number of specialists need to be brought together to perform a note-worthy investigation. Only a few fields remain that still allow lone investigators to develop meaningful insights. The most prominent form of collaboration is the co-authored publication. However, there is further potential for scientific collaboration in the form of more modular collaboration practices: academic data sharing. Here, researchers make their primary datasets available to others. This has three major benefits: first, it allows asking new research questions with existing datasets, second, it facilitates the replicability of research results, and third, it enables new research practices such as large scale meta analyses. Combined, open data in research contributes to the quantity, quality, and pace of scientific progress. Neelie Kroes, the European Commissioner for the Digital Agenda even went so far as to say, that open access to research data "will boost Europe's innovation capacity and give citizens quicker access to the benefits of scientific discoveries".
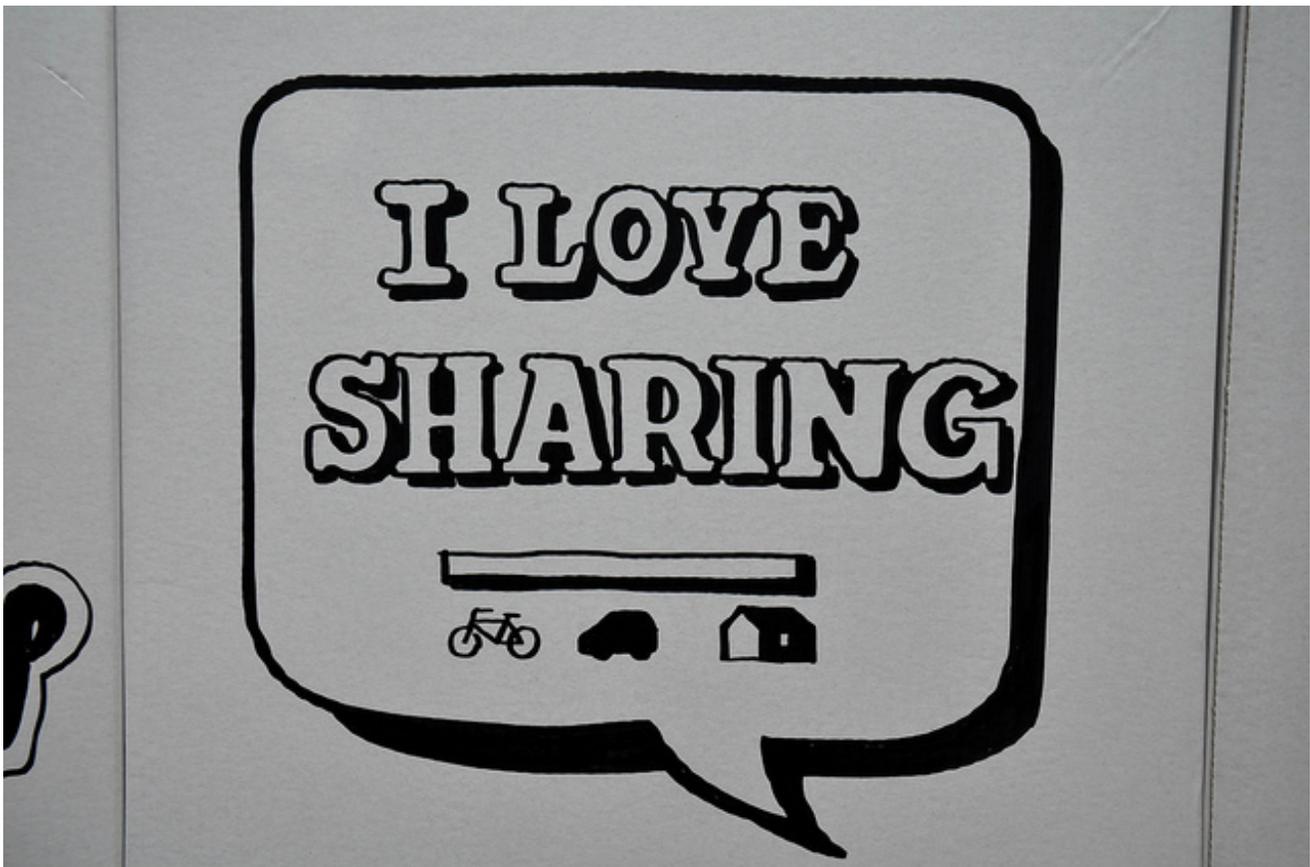
Image credit: Markus Henkel Flickr CC BY SA

Despite its advantages and prominent support data sharing sees only hesitant adoption among research professionals. In fall 2014, we conducted a survey questioning 1564 academic researchers. 83% agreed that making primary data available greatly benefits scientific progress. Yet, only 13% stated that they had published their own data in the past.

In a similar way most journals disregard the vast potential of published data. In an analysis of 141 journals from economics, Vlaeminck (2013) found that only 29 (20%) had a mandatory data sharing policy. Alsheikh-Ali et al. (2011), in an analysis of 500 research articles from the 50 journals with the highest impact factor, found that the underlying data was only available in 47 (9%) cases. In most journals publishing data is neither expected nor enforced in order to get published. This is particularly troublesome when inaccurate or wrong scientific findings are used to make political decisions — as happened in the Reinhart and Rogoff case, where false statistics justified the introduction of austerity policies. In this regard, open access to research data is not only a driver for scientific progress but also crucial for reproducibility and therefore trust in scientific results. Its meagre adoption among research professionals points to the need for new policies to motivate more academic data sharing.

Academia Is a Reputation Economy

Making data available to others is of little benefit for a researcher. Academia can be described as a reputation economy in which the individual researcher's career depends on recognition among his or her peers. The commonly accepted metrics for academic performance (the journal citation index, the Hirsch index, and even altmetrics) are all based on research article publications. Data sharing, by contrast, receives almost no recognition. As a result, researchers are geared solely towards article publications as they invest their time and resources into activities that can increase their reputation.

80% of the respondents in our survey state that the main barrier to making data available is the concern that other

researchers could published with it. At the same time, (76%) agree that researchers should generally share their data publicly. Few researchers (12%) are concerned about being criticized or falsified. These numbers show that researchers have no negative attitudes towards making data available nor are they afraid about being proven wrong. They largely recognize the potential of open access to research data. However, that does not motivate them enough to invest their time and resources into sharing their own data. This and the lack of journals that foster data sharing has led to a culture in which only a minority group, consisting of Open Access enthusiasts, publishes primary data. Today's low sharing culture reflects our academic reputation economy, in which most of one's community standing comes from article publications. We therefore believe that data sharing and reuse will only become a standard practice if it pays in form of recognition. Policies addressing data sharing need to understand academia as a reputation economy in order to work.

Why Current Policies Fail

Current policies concerning data sharing mainly fall into two camps: they either try to motivate data sharing intrinsically by invoking the common good or they force researchers to share with mandatory sharing policies. Motivating researchers to share data for the common good fails as it is not in line with the incentives of the reputation economy. Most researchers choose to invest their resources into activities that better contribute to their reputation. Consequently, debates around data sharing often focus on mandatory data sharing policies. They are embraced by funding agencies, such as the NIH in the U.S. and the Horizon2020 program in the European Union, alongside journals like Nature or PLOS ONE. Without a doubt, mandatory data sharing policies increase the number of shared datasets. However, this does not happen because researchers are motivated to do so but because it is a necessary evil to get to something else: research grants or journal publications.

And this comes with a major drawback: if data sharing is mandatory, researchers only invest the minimum time necessary to share. This in turn leads to badly labeled variables, poor documentation, and datasets that are hard to find. An empirical assessment of 18 published research papers of microarray studies showed that only 2 of them could be perfectly reproduced. In some cases it took months to reproduce a single figure.

Mandatory data sharing policies lead to a situation that makes the reuse of datasets difficult, the core reason why data sharing is advocated in the first place. To develop a culture of prolific data sharing and reuse, policy makers, funding agencies, and research organizations need to value the publication of data, it needs to pay in form of reputation.

What Appropriate Policies Could Look Like

We need a measure that indicates the importance of a dataset. Such a measure could be analogous to the citation count, which indicates the impact a research article had in the scientific community. A measure for sahred data should count publications that used a dataset (e.g., by tracking DOIs). Researchers could thus gain reputation by publishing data that gets used. And researchers could indicate their importance to a field by the number of research articles they made possible based on their published datasets.

Funding agencies should take this measure into account and privilege scientists or research groups that have a track record of distinguished datasets. By switching their policies from mandatory sharing to rewarding good datasets, funding agencies could motivate researchers not only to share but to share in a more reusable fashion.

Research communities could do more for the recognition of good datasets. Best paper awards are commonplace at conferences, in journals, and in research fields. They are welcome signs of good work that researchers use to indicate their value. Good datasets need to receive similar forms of recognition to justify the work necessary to make them publicly available in a reusable form.

And lastly, journals need to take the issue more seriously. Data journals like Nature's Scientific Data are a good first step, but need to gain impact in order to motivate the mainstream researcher to publish with them. Established

journals could instead add a data section and publish descriptions of noteworthy datasets together with their scope of application. In doing so, journals could perform the magic trick of transforming datasets into a currency researchers are used to.

Given the constant increase in complexity of many research fields, more collaboration is desperately needed. Data sharing is a form of collaboration that is worthy of our support. It is currently a desirable practice that is having a tough time gaining traction. It is like the electric car that everyone knows is good for the environment but nobody wants to buy. It is important in the current situation to set the course to promotes data sharing and rewards those who make their data easily re-usable. Only when we do this will we be able to reap the benefits that are attributed to academic data sharing.

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our Comments Policy if you have any concerns on posting a comment below.

About the Authors

Dr. Sascha Friesike is a Postdoc at the Humboldt Insitute for Internet and Society in the research area internet-enabled innovation. He holds a PhD in Technology and Innovation Management from the University of St. Gallen. Before he studied engineering economics at the TU Berlin.

Benedikt Fecher of the German Institute for Economic Research (DIW Berlin), is a doctoral researcher at the Humboldt Institute for Internet and Society. The focus of his dissertation is the participation in Open Science Projects.

Marcel Hebing works at the research data centre of the German Institute for Economic Research in Berlin. (http://www.diw.de/sixcms/detail.php?id=diw_01.c.367882.de&sprache=de)

Stephanie Linek is a researcher at the German National Library of Economics in Kiel, specializing on Science 2.0 and media psychology (http://zbw.eu/en/about-us/key-activities/usability/stephanie-linek/)