



## Data sharing shifts scientific culture

By Constanze Boettcher

Published on EuroScientist: [www.euroscientist.com](http://www.euroscientist.com)

### Not all scientists are willing to open up their data collections

Over the past few years, the scientific community has expressed concerns over the reliability of scientific research, particularly [biomedical research](#). Making the primary results of research--the actual data--more easily accessible to other scientists is seen as an important step to solve this problem. After all, reproducibility of research is at the heart of science. However, old habits die hard. And the custom of making all data fully available so that others can re-analyse and re-assess them is not yet fully ingrained in scientists' modus operandi. Training may be required to change such habits while giving credit for people producing the original data, may also encourage data sharing and enhance reproducibility.

Reproducibility is mandatory for science to functionate, says Hans Pfeiffenberger, chair of the working group on research data at research advocacy group [Science Europe](#), based in Brussels. Yet, sharing data has been encouraged by many funders. For example, Horizon2020 includes an [open data pilot](#). On a voluntary basis, researchers applying for grants in specific core research areas may indicate their [plans for making data openly available](#) as early as at proposal stage. And the ERC "strongly encourages" sharing of data with other researchers, according to its [open access guidelines](#). In the US, the National Institute of Health recommends that [preclinical research data](#) should be placed in public repositories. The US National Science Foundation also mandates that proposals outline a [data management plan](#).

Read this post online: <http://www.euroscientist.com/data-sharing-shifts-scientific-culture>

## Sharing reluctance

The trouble is that not all scientists live up to their own ideals. Scientists do regard the prospect of being able to re-analyse existing data "as the most important driver for preserving data," a study by the EU-project [Parse-Insight](#) says. But only a quarter of 1,389 survey respondents made their data openly available. Besides, in a [recent survey](#), only 13% of 1,564 mainly German researchers from various fields said they would share data publicly. Respondents to a [survey of Norwegian researchers](#) were mainly positive about sharing. Still, some remained sceptical and preferred to keep control of their data.

It is also difficult to define, what data should be shared. In a first step, data underpinning an article should be made "concurrently available in an accessible database," recommends a 2012 Royal Society report [Science as an Open Enterprise](#), which calls for making data available, so that others can access, understand and assess it. "There are absolutely no reasons in an online world not to show data if it is important enough to make a claim in a paper," says Bernd Pulverer, head of scientific publications at [EMBO](#) in Heidelberg, Germany. Already, many publishers have devised new data policies although these vary greatly, a [study](#) by researchers of the University of Nottingham, UK, says.

At EMBO publications, for example, authors may add so-called [source data](#) to papers "that is minimally processed data. Other people can extract it, re-analyse it, integrate it into their data sets and build their research from that," Pulverer says. Providing processed data supplementing journal articles has indeed become more common, Pfeiffenberger says. "But if you want research results to be reproducible, you need to provide quality assured primary data. That happens less often. And if you want to understand in great detail how someone derived the data, then you may need to have access to the raw data as well. This is extremely seldom," he says.

Researchers are uneasy about publicly sharing data for various reasons. Many scientists are concerned about [not getting appropriate credit](#) for their work. At the same time, they may consider it as problematic to spend time "on preparing data for publication in a repository or a journal, so that other researchers could use them that might compete against you," says Veerle van den Eyden, research data manager at the [UK data archive](#) at the University of Exeter. A [report](#) by the EU-project [Re-Code](#) identifies a competition for prestige and funding as a major reason why scientists do not share data openly. "The system at the moment is such that people are under huge pressure to publish a paper in a certain journal. That is what is used in research assessment tools," says Pulverer. "The usual ways of citing and indexing publications disfavour data sharing," Pfeiffenberger agrees.

Other reasons why researchers are reluctant to share data stem from lacking institutional or funder support, lacking policies as well as [lacking funding](#) for data infrastructures. Often, scientists feel [insecure about legal aspects](#), such as property rights. Moreover, not all data can be openly shared. Already the sheer amount may pose problems, Pfeiffenberger notes. Particularly in medical or social science research, [privacy concerns](#) may prevent open access to raw data. In order to make such research reproducible, "the issue is more about enabling better access to data for other experts than enabling open access," he stresses.

## Good practice in some disciplines

Still, there are disciplines where data sharing is well established, says Susan Reilly, executive director of the [Association of European Research Libraries](#) in The Hague, Netherlands. For example, in genetics or crystallography "data sharing has become the norm," says van den Eyden. She co-authored the [report](#) entitled 'Incentives and motivations for sharing research data, a researcher's perspective' that looked at successful cases of data sharing. "Researchers did their research in such a way that it could always be understood by others. It was part of the research goals. That also takes down the barriers of sharing with unknown researchers," van den Eyden says.

This culture of sharing did not appear from nowhere. In genetics, for example, policy agreements supporting data sharing have been in place for almost two decades. The so-called [Bermuda Principles](#), established in 1996, and subsequently the [Fort Lauderdale agreement](#) over prepublication data release in genomics created "a level playing field," van den Eyden says. Therefore, "data sharing could advance. Everyone knows whom the data belong to. Others

Read this post online: <http://www.euroscientist.com/data-sharing-shifts-scientific-culture>

can use them, but state the ownership," she adds. It has become common to deposit data-- for example gene sequences--"in repositories such as [GenBank](#), the [European Nucleotide Archive](#), [Dryad](#) or within the journals themselves.

Reilly recognises that there is now a move towards open data. "What has helped are the mandates," she adds. A [study](#) published in 2014 showed that data sharing policies of funders and publishers as well as new tools and infrastructure encouraged scientists to share data. Indeed, data sharing and archiving tools and platforms--such as the EMBO source data tool--need to be further developed, Pulverer urges. But he also calls for not making it "even harder" for scientists to publish. Likewise, Pfeiffenberger calls for not introducing too much bureaucracy. He regards the European policy as a good example because it allows researchers to participate voluntarily in the open data pilot.

[Preliminary statistics](#) of the DG Research and Innovation at the European Commission seem to support this view: Of 3,054 proposals so far, about 24% of applicants within the core areas opted out of the pilot, while about 27% of applicants from other fields decided to opt in. "These numbers are encouraging," Pfeiffenberger comments. He also welcomes that researchers can apply for data management resources as part of their proposals. Indeed, such financial support is necessary, Pulverer agrees. Publishers, funders and researchers have to "pull in the same direction," he says.

### Refining the sharing process

But there are also other steps that need to be taken, Reilly points out. It is "essential to increase institutional support," such as outlined in the [LERU roadmap for research data](#), she says. For example, "institutional and university libraries should become more embedded in the research process" to support researchers struggling with their data management, she holds. Training is also important. "The vision is that data sharing should be seen as a standard part of good science. This can only be achieved if you make data sharing part of research methods training," van den Eyden says.

However, recognising "not only the author of a paper but also the person who produced the data," is the key to encourage data sharing, Reilly says. Some progress has been achieved. DOIs allow proper citing of data in repositories. Data journals are another solution, Pfeiffenberger points out. Also, Thompson Reuters, which provides the commonly used indexing service web of science, has established a new [data citation index](#). But "there are still many repositories that do not offer the possibilities to cite data sets," Pfeiffenberger says.

The experts agree that the criteria for research evaluation need to be re-considered. Already, funders have begun to re-assess their criteria. "The EU now allows applicants to present data sets or software in addition to journals articles as proof of qualification," Pfeiffenberger says. The US NSF has similar requirements since 2013. But "it remains to be seen" whether reviewers really accept data sets as relevant contributions to science, he notes.

Ultimately, "data sharing will become standard practice and an indicator of high quality research. This will be not only a shift in scientific culture but also in scientific practice," Reilly holds. However, in Pfeiffenberger's view, there is no one-size-fits-all-approach. Finding a consensus within the different disciplines on how to place which type of data into the public domain, "will still take a long time," he says. He points at the importance of common journal policies. Pulverer seconds. "If everybody starts sharing data, then the playing field is level again and everybody benefits in the end," he says. Of course, "as people can reanalyse your data you open yourself up," he adds. Pulverer concludes "But this is what reproducibility is about». I strongly believe this is the scientific process."

Photo credit: [Rawpixel from Shutterstock](#)