Share your data – but not with terrorists! Why we need better policy on sharing scientific data

Scientists traditionally publish experimental findings in peer-reviewed journals. Articles generally contain a small portion of the total data produced in the course of successful and unsuccessful experiments. Recent advances in digital technologies and the internet have made it possible to make much more scientific data available more quickly, to much larger global audiences. Increasingly available data enables scientists to check if findings are correct, which in turn could increase public trust in science. More available data can also be used to investigate new scientific questions that the original scientists did not consider. Inspired by these goals, several important science funding agencies such as the Wellcome Trust in the UK and the National Institutes of Health (NIH) in the US have adopted Open Data principles to require the scientists they support to make all their data available online.

In contrast to the enthusiasm for Open Data, however, there has also been recent rises in concerns about science. Following the terrorist attacks of September 11th 2001 many governments have become concerned that research in the life sciences intended to improve our knowledge and health could also be used for malicious 'bioterrorist' purposes. Experiments to understand the genetics and functioning of viruses that affect humans such as smallpox and polio have been identified as examples of research that could have such 'dual use' applications. The international security community has argued that scientists have a duty to identify research with harmful potential and take extra care not to publish or make available data that could be misused in this way. Security concerns are increasingly making their way into daily scientific life through concerns about the misuse of *published research.*

These two areas of concern present scientists with two contrasting sets of obligations: to share as much data as possible, and to be careful about the data that they publish. While such obligations are not automatically contradictory, the manner in which they area represented to scientists is often highly problematic. Dual use and Open Data policies are written by different groups who rarely, if ever, meet. The duties that they present to scientists rarely reference the other set of obligations. Moreover, they present different accounts of both what 'data' is and what scientific responsibility looks like.

To produce a coherent 'joined up' policy on how scientists should responsibly manage their data these groups need to work together. A new framework is needed that recognises that scientists draw on multiple different types of information from a range of different sources in any one experiment or set of experiments. In addition data can be made available through an increasing variety of online platforms from databases to peer-to-peer file sharing systems. Data may not only be shared or sought at the end of an experiment, but at many points during the life cycle of a particular piece of research.

A responsible data management policy must focus on these small every-day transactions, which we term ‘micro-movements’ of data. The duties of each individual scientist searching for, or sharing, data are i) to assess where the data are coming from and consider whether the source raises any concerns, ii) to consider where their data are going to, and whether this choice increases the potential for desirable reuse of the data, and iii) to think about what safeguards are in place for anywhere that data are stored and to choose a storage site with appropriate protections for the types of data being deposited. Future responsible data management guidelines should focus on helping scientists to make these judgements about everyday data use and reuse.

Bezuidenhout LM[1,2], ***Morrison M***[3]
[1]*Steve Biko Centre for Bioethics, Faculty of Health Sciences, University of the Witwatersrand, Parktown Johannesburg, South Africa*
[2]*Egenis Centre for the Study of the Life Sciences, University of Exeter, Byrne House St German's Road, Exeter, Devon, United Kingdom*
[3]*Centre for Health, Law and Emerging Technologies (HeLEX), Nuffield Department of Population Health, University of Oxford, Ewert House Banbury Road, Oxford, United Kingdom*

Publication

Between Scylla and Charybdis: reconciling competing data management demands in the life sciences.
Bezuidenhout LM, Morrison M
BMC Med Ethics. 2016 May 17