Making research data and statistical analysis commands accessible to other researchers has important advantages of transparency, accountability, replicability, and efficiency, as was made clear in the commentary by West [1]. In addition, broad availability of research data can help to mobilize the world’s intellectual community to better harvest the scientific potential of the exponentially increasing number and size of data and databases. Indeed, many researchers and institutions produce and store substantially more data than they can effectively explore and analyze in all possibly relevant and promising respects. Reasons for this are that time and resources are limited, but also that collected data are often a source of knowledge that goes beyond the locally available creativity and expertise. Recognizing this has been decisive for the success of international research collaborations such as the Human Genome Project [2–4]. Also, the sharing of data from large, longstanding cohort studies [5] and digitalized patient records [6] and the emergence of individual patient data analysis [7] are examples of using original data more broadly. The Institute of Medicine has recently elaborated guidance on how to responsibly share clinical trial data [8,9]. Apart from intended multicenter collaborations to exchange or link data and join efforts, open access of research data can evoke unexpected innovative creativity in using them. This can promote the expansion of knowledge for scientific purposes and public interests and avoid that important data grow too old or will never be harvested and therefore wasted.

However, in the face of this attractive perspective for the public good, a number of issues have to be addressed if full data are to be published together with research reports or made publicly available in another way.

As also mentioned by West [1], researchers, institutions, and funders will want to know how intellectual property will be respected. Will the efforts to develop an original study, get funding, and collect data be acknowledged as authentic steps in making an academic career? Would the primary researchers agree with enabling others, possibly their academic competitors, to freely use the data for own purposes? Could this conflict with productive academic competition and discourage collection of original data when these become immediately available to anyone else? Would this stimulate primary researchers to wait very long with publication of their data, until they feel safe that no one else could harvest new findings before they have done this? And can we expect a significant increase of cumbersome negotiations on multinational mega-authorships [10,11]. A related issue is to what extent the primary researchers should decide on what additional analyses could be done. If their involvement would not be compatible with allowing other researchers to have a fresh look at the data, it might be better if an independent third party would look after this.

As empirical research costs a lot of money, it is reasonable, especially if there is no collaboration mutually serving the scientific interests of both the primary researchers and secondary users, that the latter contribute to the costs proportionally. Maybe we also need funding methods and practices that are more focused on data sharing. Funding regulations can adopt the condition that data, after being used for the purpose as funded, should be handed over to a platform that is supported by public and possibly also private funds.

A further question is whether a distinction should be made between various types of users and whether some of these should be excluded from use of freely available databases. Should, for example, industries be refused to get access for commercial interests [1] or would it be better to see their involvement as efficient use of patient data compensated for by asking an extra high price? If so, should this be managed by the primary researchers or by a public authority? Or should products be cheaper if manufacturers freely get research data which they could otherwise only obtain by funding research?

There are also challenges in ensuring data quality and in providing methodologically solid guidelines for data use, interpretation, and transparent reporting. Moreover, we must safeguard the privacy of patients and appropriately

Conflict of interest: The author report no conflicts of interest. The author alone is responsible for the content and writing of the article.
* Corresponding author. Tel.: +31 43 388 2302.
E-mail address: andre.knottnerus@maastrichtuniversity.nl

http://dx.doi.org/10.1016/j.jclinepi.2015.05.034
0895-4356/© 2016 Elsevier Inc. All rights reserved.
deal with informed consent, also exploring innovative consent approaches that serve the public interest without negatively affecting the individual. In addition, we have to counter scientific misconduct and abuse of publicly available data for destructive purposes [12]. Furthermore, we must keep in mind that secondary analysis should not be an unprepared surfing or fishing expedition [13], but—like all research—requires a valid protocol that has survived independent review. These issues must be well addressed by responsible investigators and covered by research boards and ethics committees—also in the context of international collaboration.

One might think that science cannot develop without competition, focused on getting and staying ahead with collecting own data and put them under lock and key. Others may be unhappy with emphasizing scientific competition and strive for common responsibility of researchers internationally. But there are ways to combine productive competition with stimulating data sharing, for example, by acknowledging project initiators and data collectors by linking them routinely to the data and by citing their basic work. Moreover, it should be a significant indicator of esteem for researchers if the data they have gathered are used worldwide to serve science and society at large. Moreover, articles publicly presenting the original data used can be signalized with a quality mark [1]. In addition, analogous to trial registries, data sources can be registered and linked to those who started them. This would also enable oversight over analyses being done and research reports produced. As to the question who should manage such a registry, we can learn from trial registries and the Cochrane Collaboration [14], with their great performance in information sharing, disseminating, and transparency.

These considerations make clear that much additional thinking and exchange of experience are needed on how to optimally deal with the both promising and complex development of international data sharing. At the same time, these challenges should not discourage progress in realizing the opportunities and gains of intensified knowledge harvesting. We seem to move toward an era that needs a new paradigm, that is, a more open, public view on intellectual property. This is legitimate, as knowledge comes from data provided by independent research subjects and populations, and is financed—directly or indirectly—by society. Knowledge, and the data it comes from, can also be considered as a common-pool resource, [15] a common property resource, or a global public good [16]. In such a perspective, we should get away from everlasting possession of data by individual researchers, institutions, or companies. This is also a matter of well-perceived own interest, as in an era of emergence of Big data, researchers cannot stay purely individualistic. The more possessive they will stay, the less progress they will achieve. This is also what we learned as solo researchers disappeared from the laboratories and from international epidemiologic research.

Speaking about international data sharing, there is a clear analogy with the open access publishing movement [7], also in its seeking for new business models that fit with maximal accessibility. This, and the emergence of the Cochrane Collaboration and trial registries, shows that there are good reasons to rely on the self-regulating scientific community, with its checks and balances in moving forward. In this context, also editors and their networks have their role, as they had in promoting adherence to reporting standards and trial registries, for example, by already encouraging authors to provide their data for replication purposes and to contribute even more to the public good.

References


