

Anonymous Reviews:

FACILITATING DATA SHARING IN THE BEHAVIOURAL SCIENCES

REVIEWER 1:

In this article, the authors suggest that one method for increasing awareness of the benefits of data sharing can be affected by administering a survey, hopefully encouraging either a minority or majority influence on other members of the behavioural science community. The authors describe the process for creating such a survey and provide their survey, although untested. This is a well-timed article as the existing literature on social and behavioural science data is quite thin. The majority/minority influence of norms is a nice theoretical construct to explore. This is definitely one approach to increase awareness of the benefits of data sharing.

That said, this article needs to address a number of underlying assumptions, specifically with use of terms that are not as stable as the authors use them. In order to make their argument more clear, the authors need to more fully address (i) the data in question, (ii) data sharing, and (iii) data use. What is meant by “data” differs greatly by community. It would behove the authors to expand on the data they foresee being shared and the role of this data in research. Data sharing is a similarly contentious term. What do the authors mean by data sharing? What is entailed in sharing? What level of access or support? For example, by data sharing do they mean available online from a researcher’s webpage with the assumption that a potential user would contact the researcher for more information about the data context or does the data have to be deposited in a community-sanctioned repository with complete metadata? Would the data be used if it were shared? For instance the World Values Survey that is brought up in section 2, what are the usage statistics for this dataset? And even further what sorts of uses are being counted in the usage statistics: data downloads, publications from the data, etc.? By addressing these issues, the article will be more clearly scoped and support generalization beyond the behavioural sciences.

Minority and majority influences need to be placed in relationship to other influences. The behavioural norms do not exist in a vacuum – in addition to practices there are incentive structures and technological barriers to data sharing. The authors are clearly addressing the practices, but without also stepping back and addressing the infrastructure within which these practices are located, otherwise changing the practices will be for naught. Affecting change through the use of a survey seems a bit pie-in-the-sky, but in concert with other efforts this would be very strong.

Please note that ethics boards do not outright prohibit data sharing (or at least this is true of the IRB in the US), they merely ask you to disclose what you plan to do with the data. If you inform participants that their data will be shared outside of the current research, the ethics boards will not outright halt a project unless they think this will adversely affect the participants. Oral histories are a good model for social science and behavioural research, as they are collected with the intention of making them available in the future, sometimes without an embargo period.

As to the implementation of this article, it is remarkably short (a full 7 pages), through more

judicious use of citations (4.5 pages of references is about 3 too many), the authors could really push this paper out a couple more pages. The introduction is quite general, and does not set up the paper very strongly. For instance, by the time I got to the section on the Survey Research Instrument (section 3.3), I was unclear as to whether the SRI was being setup as a means for studying the problem or affecting change, or both. The authors need to provide more guidance as to how the reader should interpret the role of the SRI. The treatment of norms (section 3.1) is unnecessarily long; a dictionary definition would serve better than the clothing metaphor. There are a few awkward sentences, such as the last sentence of the 3rd intro paragraph and the middle of the 3rd paragraph in the 2nd section, and the entire paper could use an editing pass. There is very little treatment of existing data sharing literature, see references below. A fuller treatment of this topic will help alleviate some of the assumption questions above.

This article has much potential, and I look forward to being able to cite it in my own research, pending changes.

Boruch, R. F. (1985). Definitions, products, distinctions in data sharing. In Fienberg, S. E., Martin, M. E. & Straf, M. L. (Eds.). *Sharing Research Data*. Washington, DC, National Academy Press: 89-122. Retrieved from http://books.nap.edu/catalog.php?record_id=2033 on 29 December 2009.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A. & Blumenthal, D. (2002). Data withholding in academic genetics: Evidence from a national survey. *Journal of the American Medical Association*, 287(4): 473-480. Retrieved from <http://jama.ama-assn.org/cgi/content/full/287/4/473> on 29 December 2009.

Chin, G. J. & Lansing, C. S. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. Conference on Computer Supported Cooperative Work, Chicago, IL, Association for Computing Machinery. Retrieved from <http://portal.acm.org/citation.cfm?id=1031677> on 29 December 2009.

Chubb, J. M., Austin, E. W., Geda, C. L. & Traugott, M. W. (1985). Sharing research data in the social sciences. In Fienberg, S. E., Martin, M. E. & Straf, M. L. (Eds.). *Sharing Research Data*. Washington, DC, National Academy Press: 39-88. Retrieved from http://books.nap.edu/catalog.php?record_id=2033 on 29 December 2009.

Fienberg, S. E., Martin, M. E. & Straf, M. L. (Eds.). (1985). *Sharing Research Data*. Washington, DC: National Academy Press. Retrieved from http://books.nap.edu/catalog.php?record_id=2033 on 30 December 2009.

Hedrick, T. E. (1985). Justifications for obstacles to data sharing. In Fienberg, S. E., Martin, M. E. & Straf, M. L. (Eds.). *Sharing Research Data*. Washington, DC, National Academy Press: 123-147. Retrieved from http://books.nap.edu/catalog.php?record_id=2033 on 29 December 2009.

Hilgartner, S. & Brandt-Rauf, S. I. (1994). Data access, ownership and control: Toward empirical studies of access practices. *Knowledge*, 15: 355-372.

Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. (2003). Meeting organized by the Wellcome Trust, Fort Lauderdale, Florida, Wellcome Trust. Retrieved from

www.wellcome.ac.uk/.../groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf on 29 December 2009.

REVIEWER 2:

This is an interesting article that has proposed a novel solution to the problem of encouraging data sharing in the behavioral sciences. The paper suggests that norms around data sharing can be changed with the use of a survey designed to change opinions and intentions. By using the foot-in-the-door technique, a survey could gradually introduce data sharing ideas and could lead to greater willingness and intention to share data. With enough people taking such a survey, norms in the behavioral sciences could be changed. In the end, however, I am sceptical about the claims made in this paper, it is not clear that it makes an appreciable contribution, and I have serious concerns about the ethics of what is proposed here.

The paper provides a strong literature review of both the research around surveys as tools for norm changes and around the challenges and benefits of data sharing. One concern that shows up in the literature review but runs throughout is that the paper focuses almost entirely on the benefits of data sharing without addressing any of the challenges or potential downsides. Sharing data is difficult and potentially risky for scientists. Making datasets useful outside of the original context takes additional and often uncompensated work [1,5]. When scientists share data, they open up potential risks, including being scooped on discoveries or having mistakes pointed out [2]. Others have found that resistance to data sharing is not a matter of norms, but of lack of demand, concerns about data ownership and quality, etc. [3]. While it can be argued that sharing is good for science and society writ large, the authors would do well to recognize that the individual benefits may not be aligned with the social benefits.

This leads into another aspect of the paper that I find problematic: the paper implies that the low level of data sharing in the behavioral sciences is primarily an issue of norms and attitudes. As mentioned above, there is a body of research that attributes resistance to data sharing to other factors. Recognition structures for scientists (tenure & promotion, publication, grant funding) tend not to provide adequate rewards to incentivize data sharing. There is frequently no financial or human resource support for data sharing. Even if scientists support the idea of data sharing, there are many other factors (and other norms) that may conflict, including norms about how to get ahead in science, norms about data ethics and data privacy, etc. In the conclusion, the authors say that they have provided suggestions to “facilitate” data sharing, but I would argue that these suggestions merely *promote* data sharing, and do little to make it easier to do. I do not mean to suggest that norms aren’t part of the picture or that attempts to directly change norms would not be useful; but these other factors provide important context and framing for the data sharing issue and should be addressed in the paper.

The discussion of norms, behaviors and attitudes presents a confusing picture of the relationships among these concepts. Sometimes (as in the discussion of the non-sharing norm in section 2), a norm seems to simply be collective behavior. Thus, because most people don’t share, not sharing is the norm (regardless of how people think or feel about sharing). In that section, attitudes are distinct from norms (c.f. “the extent to which *attitudes and norms* lead to data sharing problems”). Later in section 3.1, however, “norms are by definition the attitudes of the majority.” This vagueness about these definitions makes it difficult to understand and assess how the survey is expected to work – is its goal to change attitudes, or to change behaviors, or both?

A particularly weak point in the paper is the discussion of the development of the survey instrument in section 4.1. It is not clear how the principles discussed were applied to this specific example, or which of the questions in the survey are specifically designed to promote

data sharing norms. There are also claims in this section that run counter to the literature. For example, it is proposed that a small request to use as a starting point would be to ask scientists if they would be willing to reuse other scientists' data: "It is very unlikely that people would refuse to use data that they themselves did not collect." Prior research has found the opposite: scientists tend to be very wary about using data that they did not collect themselves. They have significant concerns about the skill and trustworthiness of the scientists who originally collected the data, and need to know that the dataset is understandable and relevant to and compatible with their questions and approach [4].

In section 4.2, there is a suggestion that giving the survey at workshops would be a way to reach a large number of people, but this seems to be an example of "preaching to the choir." One would expect that people who attend such workshops would be the ones most likely to already see data sharing as valuable and already engaged in discussions about the topic.

This paper needs significant editing for spelling, grammar and clarity.

I am deeply troubled by the lack of a discussion of the ethics of doing what is suggested in this paper. This technique of using a survey to change an opinion or behavior is presented here in sterile academic language, but it is deeply controversial. First, the practice relies on deception: it is doubtful that the survey would be effective if the purpose were disclosed, for example, by a statement at the beginning that said, "The purpose of this survey is not to find out how you would answer these questions, but instead to subtly influence your attitudes and behaviors." In the political realm, this same technique is known as "push polling," and is generally considered to be unethical (c.f. <http://www.theapc.org/default.asp?contentID=552>). This kind of subliminal attitude adjustment may be slightly more acceptable in fields like public health (where there is significant individual benefit to quitting smoking or eating well), or in political participation (where there is little if any risk), but as discussed earlier, despite the positive picture presented here, data sharing is not universally valued and is not without individual risk to the scientists and their careers. Second, while this technique may increase data sharing, it may also harm the academic enterprise by contributing to the skepticism and survey fatigue that is leading to lower response rates and lower-quality data on surveys generally.

Finally, at the end of the paper, it is unclear what contribution it makes. The concept of changing norms and attitudes using surveys is not new, and this paper does not add anything novel to that conversation. The paper does not change the way we understand the value of data sharing generally or specifically in this domain. The idea of using a survey to influence data sharing is a) relatively trivial, and b) not well thought-out in terms of context-specific issues (e.g. norm conflict, structural barriers to data sharing) or ramifications (ethics). In order for the survey itself to be a contribution, I would expect it to be tested and validated in some form to show that it has the predicted effects. In the end, this feels more like a research proposal than a finished article.

1. Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation*, 4(2), 12-27.
2. Birnholtz, J., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. In *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 339-348). New York, NY: ACM Press.
3. Borgman, C., Wallis, J., & Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal*

- on Digital Libraries*, 7(1), 17-30.
4. Faniel, I., & Jacobsen, T. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW)*, 19(3), 355-375.
 5. Zimmerman, A. (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7(1), 5-16.