Concerns of Open Data

Open Data hasn’t been fully embraced by everyone. Before these people commit to Open Data, they see problems that need to be solved first. These problems create barriers to making their data public. Researchers cite legal, practical, and competitive concerns that have kept them from advocating and acting for Open Data.

Legal concerns that cause apprehension for some are about rights and data confidentiality. From the findings in the Tenopir et al. (2011) paper, one of the main concerns of researchers is that they don’t have the rights to make the data public (Houtkoop et al., 2018, pg. 77). Part of this problem can come from sponsored research. Sponsors can require confidentiality in the study, meaning they don’t want the data to be available to others (Leetaru, 2017). Funding is integral in conducting research, and that funding can come from private sponsors trying to use the results for future products. Sponsors don’t want to potentially miss out on profitable results because someone else uses the data first to do so. Researchers don’t want to not be able to conduct the research they want to, so they could be forced to keep the data private since they don’t have the rights to make it public. There is also data confidentiality regarding participants and those involved in the study that cause concern. Anonymity is crucial in conducting sensitive research, and some worry that making the data public would hurt those who participated.

In sensitive research studies, there is a risk that the participants could be identified, but there are ways to decrease the risk and secure confidentiality. The most common way to protect identity is removing direct identifiers provided in the datasets (Fraser & Willison, pg. 13, 2009). K-anonymity is used to protect against identifying someone by indirect variables. It involves making the variable into a larger category, such as age to age group or birthdate into birth year (Fraser & Willison, pg. 13, 2009). Pseudonymization is used to create a fake “ID” for a participant across multiple datasets (Fraser & Willison, pg. 15, 2009). Besides direct identifiers, listing specific dates could lead to that person being identified. Many people have access to birthdays and those closer to the individual have a chance of knowing when surgeries or treatments took place. To help protect this individual, specific dates like birthdays or surgeries should be listed in “durations or intervals” (Fraser & Willison, pg. 20, 2009).

Practical concerns that create barriers for some are about the process itself of making data public. Money, time, set standards are seen as three practical complications to pursuing Open Data. In the paper from Tenopir et al. (2011), the two biggest concerns cited by researchers were lack of funding and inefficient time and one of the most common minor concerns were researchers not having set standards for making data public (Houtkoop et al., 2018, pg. 73). Lack of funding leads to researchers fears of wasting time and money because of lack of set standards. For researchers who are new to Open Data, it might be hard finding out all the intricacies of it, such as where and how to make their data public. When researchers don’t know the data sharing standards that means in order to make their data public, they must find out how and where to publish data and then actually publish it. This process could mean less time and money being devoted to aspects of the study that may be required or mandatory.

However, there are set standards created by The Center for Open Science, called the Transparency and Openness Promotion. There are three levels that a journal can adopt into their submission guidelines that researchers can follow and know what and how to utilize data sharing (TOP Guidelines). As more journals start to adapt to the Open Science movement, more researchers will know how to share their data just by going to submit their work, instead of having to look it up specially.

People also are concerned about Open Data due to the competitive nature of conducting and publishing research. Competitive barriers are ways in which making data public would disadvantage the original researcher. Loss of credit is an issue, where researchers don’t want to make their data public before they publish their own results because they don’t want their data “scooped”, or the data is published by another researcher before your own is (Houtkoop et al., 2018, pg. 73- 77). With making data public, researchers may also feel that a loss of control over their own work is a competitive barrier to Open Data. For example, researchers asked in one study said they were concerned about others rejecting their own conclusions because they analyzed the data differently, others findings errors in the data, others misinterpreting the data, and “loss of control over intellectual property” (Houtkoop et al., 2018, pg. 77). Some studies can result in data that can become “patentable intellectual property” but making the data public can hurt the patent process (Barron, 2018). In that case, making the data public can result in both a loss of control and a loss of potential money and funding from sponsors.

Open Data isn’t the end of marketable research. One way for researchers to maintain their competitive edge is requiring secondary researchers to cite the original researchers when using their data (Houtkoop et al., 2018, pg. 73). The fear of “scooping” has also been addressed. In one study, two research labs who both used open science practices were observed as they went through a research project. It found that their fears of “scooping” were reduced when they focused on the research itself, as opposed to publishing being the main motivator (Laine, pg. 12, 2017).

References

Barron, D. (2018, August 13). How freely should scientists share their data? *Scientific American: Observations*. Retrieved from https://blogs.scientificamerican.com/observations/how- freely-should-scientists-share-their-data/

Fraser, R., & Willison, D. (2009). Tools for de-identification of personal health information. Pan Canadian Health Information Privacy (HIP) Group.  
 http://www.ehealthinformation.ca/wp-content/uploads/2014/08/2009-Tools-for-De- Identification-of-Personal-Health.pdf

Authors (Year). Guidelines for Transparency and Openness Promotion (TOP) in journal policies and practices: “The TOP Guidelines”. Retrieved from https://osf.io/ud578/

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1,* 70- 85. <https://doi.org/10.1177/2515245917751886>

Laine, H. (2017). Afraid of scooping- Case study on researcher strategies against fear of scooping in the context of open science. *Data Sharing Journal*, 16, 1-14

Leetaru, K. (2017, July 20). Should open access and open data come with open ethics? *Forbes.* Retrieved from <https://www.forbes.com/sites/kalevleetaru/2017/07/20/should-open-> access-and-open-data-come-with-open-ethics/#1524d65a5426