

# Alternatives to Social Science One

Margaret Levi, *Stanford University*

Betsy Rajala, *Stanford University*

## ABSTRACT

This article responds to King and Persily's (2019) proposal for a new model of industry–academic partnership using an independent third party to mediate between firms and academics. We believe this is a reasonable proposal for highly sensitive individual-level data, but it may not be appropriate for all types of data. We explore alternative options to their proposal, including Administrative Data Research Facilities, Data Collaboratives at GovLab, and Tech Data for Social Good Initiative at the Center for Advanced Study in the Behavioral Sciences. We believe social scientists should continue to explore, evaluate, and scale a variety of industry–academic data-sharing models.

Private companies possess valuable data that are largely inaccessible for social science. The incentives for academics and industry are sufficiently different to make any scalable collaboration difficult. King and Persily (2019) offer a solution. They propose a partnership model that is based on an independent third party (i.e., Social Science One) that adjudicates between companies and academics on issues of data distribution. This is ideal for collaborations for which protecting the security of fine-grained individual-level data and a propriety underlying algorithm is a necessary condition for making the data available to academics. A third-party adjudicator is essential when a company confronts external pressures to release data regarding something the world is desperate to understand. Social Science One is attempting to leverage its mediation role to produce a mutually beneficial agreement that ensures data privacy and addresses a company's reputational concerns but that also prioritizes data quality. Facebook's data could reveal important new forms of political influence, and King and Persily are working to ensure that the data are made available and analyzed responsibly.

Of course, this is not a one-size-fits-all model, nor is it intended to be. There is a variety of data that, although not collected with scholarly research purposes in mind, turn out to be useful as evidence in academic claims. For example, Putnam (2000) repurposed marketing data for his book *Bowling Alone* to show how individuals have become increasingly disconnected from their family, community, and democratic structures.<sup>1</sup> In economics, Cohen et al. (2016), Cook et al. (2018), and Cramer and Krueger (2016) used Uber data to explore questions of consumer surplus, the gender gap, and how technology has changed the transportation industry, respectively. Although these examples are not as institutionalized

as Social Science One, they do provide different types of collaborations that previously worked and could be replicated.

These types of partnerships drastically reduce the costs for both academics and industry. Researchers are free to explore questions that can be answered with the data provided; they do not need to go through the process of submitting an extensive proposal to a third party, and companies provide only the data that fit with their business interests. Of course, this may mean that researchers are not granted access to all of the data that they want. However, attempts to access a company's entire data archive should not delay or prevent access to some of its data. In many cases, even partial data from private companies can surpass the quality of alternative data sources. Having access to entire datasets from a wide variety of companies is the ideal, of course, but it simply is not realistic—yet.

To obtain access to the largest possible proportion of private data, social scientists must use a variety of different partnership models. Fortunately, there are several efforts exploring additional data-sharing models for academic–industry partnerships, including the following:

- *Administrative Data Research Facilities (ADRFs)* collate government and private data across agencies, companies, and jurisdictions in a secure yet accessible way.<sup>2</sup> ADRFs act as both a data-storage facility and an intermediary to assess the validity of research questions. However, the adjudication function of ADRFs is not as intensive as Social Science One. Therefore, this model of collaboration will meet the needs of a wide range of data producers except those that have serious reputational concerns requiring a more hands-on approach to determine acceptable research questions (e.g., Social Science One's relationship with Facebook).
- *Data Collaboratives at GovLab* allows partner firms to engage in various approaches ranging from reliance on trusted intermediaries, in the spirit of Social Science One, to the creation of data cooperatives, in which data are provided to one organization or researcher.<sup>3</sup> This option is

Margaret Levi  is the Sara Miller McCune director at the Center for Advanced Study in the Behavioral Sciences and professor of political science at Stanford University. She can be reached at [mlevi@stanford.edu](mailto:mlevi@stanford.edu).

Betsy Rajala  is a program director at the Center for Advanced Study in the Behavioral Sciences at Stanford University. She can be reached at [betsy.rajala@stanford.edu](mailto:betsy.rajala@stanford.edu).

---

designed for organizations that want to co-create a case-by-case collaborative designed to fit the needs of a company regarding a specific piece of data. It does require companies to actively engage in the design process.

- *Tech Data for the Social Good Initiative at the Center for Advanced Study in the Behavioral Sciences (CASBS)* focuses on making aggregate or archived datasets publicly available to academics.<sup>4</sup> In this case, the expectation regarding company involvement is limited: companies provide only the data that they are comfortable making available to any and all researchers. Like ADRFs, this model is not designed for the extensive engagement of the data producers regarding which questions or researchers can gain access.

All of these models are ongoing efforts that continuously evolve in response to the successes and failures of previous partnerships. In fact, most of them are so new that it is unclear exactly which conditions will lead each to succeed or fail. Like Social Science One, they all attempt to align interests between academics and industry, sometimes by avoiding sensitive topics or selecting questions in which both are interested. All are important experiments, but none are perfect. However, as a group, they provide researchers with a starting point to determine the ideal collaboration model for a given situation.

All of these efforts, including Social Science One, are both novel and experimental. Evaluation of which is best suited for what type of data and circumstances is still in the future. Exploration of diverse forms of cooperation is the first step; second is the documentation of what works and what does not, including discovering and ensuring benefits to all partners. With time and analysis, we can begin to understand the conditions that foster trust relations between independent researchers and industry. Protocols and rules that guard the interests of each party ultimately should facilitate greater willingness by all participants to devise even more expansive data-sharing arrangements to foster the use of private data to advance scientific research.

King and Persily (2019) provide a reasonable approach to academic–industry partnerships for highly sensitive data and proprietary information, particularly when a company has major reputational concerns. However, private companies own a wide variety of different types of data useful to academics that will require different types of collaborations. Partnerships seldom require the degree of oversight modeled in Social Science One. Given the relatively early stage of these partnerships, additional data-sharing models should be explored, evaluated, and scaled until we have a set of effective partnership models for all types of data. ■

---

#### NOTES

1. The marketing data used were from the advertising firm DDB Worldwide of Chicago. Available at [http://bowlingalone.com/?page\\_id=7](http://bowlingalone.com/?page_id=7).
2. Additional information about ADRFs is available at [www.adrf.upenn.edu/urba\\_nadrf](http://www.adrf.upenn.edu/urba_nadrf).
3. Additional information about Data Collaboratives at GovLab is available at <https://datacollaboratives.org>.
4. Both authors are currently involved in this project. Additional information is available at <https://casbs.stanford.edu/programs/projects/tech-data-social-good-towards-public-facing-tech-data-causal-analysis>.

---

#### REFERENCES

- Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. 2016. "Using Big Data to Estimate Consumer Surplus: The Case of Uber." Cambridge, MA: National Bureau of Economic Research. *NBER Working Paper No. 22627*.
- Cook, Cody, Rebecca Diamond, Jonathan Hall, John A. List, and Paul Oyer. 2018. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers." Cambridge, MA: National Bureau of Economic Research. *NBER Working Paper 24732*.
- Cramer, Judd, and Alan B. Krueger. 2016. "Disruptive Change in the Taxi Business: The Case of Uber." *American Economic Review* 106 (5): 177–82.
- King, Gary, and Nathaniel Persily. 2019. "A New Model for Industry–Academic Partnerships." *PS: Political Science & Politics*. Available at doi:10.1017/S1049096519001021.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.