A New Model for Industry-Academic Partnerships[1]

Gary King[2] and Nate Persily[3]

4/9/2018

Abstract

The mission of the academic social sciences is to understand and ameliorate society's greatest challenges. The data held by private companies holds vast potential to further this mission. Yet, because of its interaction with highly politicized issues, customer privacy, proprietary content, and differing goals of firms and academics, these data are often inaccessible to university researchers. We propose here a new model for industry-academic partnerships that addresses these problems via a novel organizational structure: Respected scholars form a commission which, as a trusted third party, receives access to all relevant firm information and systems, and then recruits independent academics to do research in specific areas following standard peer review protocols organized and funded by nonprofit foundations. We also report on a partnership we helped forge under this model to make data available about the extremely visible and highly politicized issues surrounding the impact of social media on elections and democracy. In our partnership, Facebook will provide privacy-preserving data and access; seven major politically and substantively diverse nonprofit foundations will fund the research; and the Social Science Research Council will oversee the peer review process for funding and data access.

[1] Draft in progress; comments welcome (send to King@Harvard.edu). The current version of this paper is available at GaryKing.org/partnerships.

[2] Gary King, Albert J. Weatherhead III University Professor, and Director Institute for Quantitative Social Science, Harvard University (1737 Cambridge Street, Cambridge MA 02138), GaryKing.org, King@Harvard.edu.

[3] Nate Persily, James B. McClatchy Professor of Law, Stanford Law School (559 Nathan Abbott Way, Stanford CA 94305-8610), www.persily.com, npersily@stanford.edu.

1. Introduction

To deliver and improve their popular products, modern internet technology firms collect large quantities of data about human behavior. This information enables companies to make better decisions by building on research and methods from the social sciences and other fields. However, it also holds the potential to enable far more scientific discovery, advance social good, and provide the insights to understand and ameliorate important problems afflicting human societies. Successful industry-academic collaboration can simultaneously help a company, advance science, and potentially solve some of the society's most pressing challenges.

Yet, progress in and from data sharing for social good will only occur if individual privacy is protected, company trade secrets and related proprietary information are respected, and the standards and independence of the scientific process are secured. Even though "academics provide the creative fuel for much early-stage research that leads to industrial innovation," achieving these goals simultaneously is difficult in any area (Jasny et al., 2017), but it has proven especially difficult for the novel data types collected, and often invented, by internet technology firms.

We propose here a new type of organizational structure for industry-academic partnerships designed specifically to span the divide between the needs of internet technology firms and academic researchers in highly politicized environments, so that ultimately all parties, including society at large, will benefit. We begin by describing the problem we seek to solve and our proposed model. We then describe the success we have had so far in applying this model to study the impact of social media on elections with a partnership we forged for academics with Facebook and eight major diverse nonprofit foundations. We then give some principles by which partnerships we have designed should follow.

2. The Problem

For most of their history, the academic social sciences collected or purchased their own data and so have had only occasional need for formal data use agreements and other relationships with industry. When they had more of a need, they followed the traditional model for industry-academic partnerships in the natural and physical sciences where private firms donate funds, data, or expertise to a university lab for a specific research project or program, often in return for considerations such as right of first refusal to license any resulting patents and pre-publication review (but not pre-approval) (e.g., Corzo and Eastman, 2015). This model works well where a company funds, and perhaps even details a few employees to work at, a

university lab for a specified period of time. The academic researchers then operate independently, have no pre-publication approval by the firm, and have unfettered control over their research agenda and methodological choices. This time honored partnership model has generated enormous value for academic researchers, private firms, the scientific community, and society at large.

Unfortunately, the difficulties of collaboration with academia in the era of big data about human behavior means that this traditional industry-academic partnership model does not work in many areas. Not long ago much social science research could be completed without industry, since most data was created by academics or accessible from governments or firms making data public. Today, big data collected by firms about individuals and human societies is more informative than ever before, which means it has increasing scientific value but also more potential to violate individual privacy or help competitors. Although many types of social science research require a partnership with private firms even to begin a study, many firms are understandably more leery of sharing data. In other words, social scientists have access to more data than they ever had before to study human society, but a far smaller proportion than at any time in history.

The problem we must overcome is that neither of the two logical antipodean approaches work to solve the problem -- especially for high profile, highly politicized, sensitive issues. In the first approach, academic researchers would remain fully independent, without pre-publication approval. Unfortunately, even if a large tech firm were willing to share data with many researchers and the privacy of individuals could be assured, proper inferences require the full chain of evidence from the world to the data, in this case often requiring proprietary information about the firm's policies, practices, procedures, and platforms, and sometimes even access to its computer systems. Published scholarly articles accessing proprietary data in this way without legal agreements requiring nondisclosure and pre-publication approval are rare, although a variety of clever ad hoc compromises and approximations have been used in specific situations (e.g., Chen et al., 2017).

In the second approach, academic researchers sometimes go inside companies and become consultants. They sign nondisclosure and other legal agreements and obtain all necessary data, information, and systems necessary to do novel research, but they have at least somewhat limited ability to publish freely. We can think of these academic consulting arrangements as falling on a continuum: At one end, they can be highly restrictive, with tight control and pre-publication approval -- such as for sensitive issues close to the firm's core products. At the other end, they can have lax, or pro-forma regulation -- usually when firms benefit from openness, allow collaborations between internal and external researchers, can patent before publication, and allow research on topics orthogonal to core products.

Consultant contributions to the firm can be large and may be of the highest scientific quality, but their contributions to the scientific community become complicated as we move closer to the first end of the continuum. In addition to the well known effects of financial and other conflicts of interest on research conclusions (Banaji and Greenwald, 2013; Wilson and Brekke 1994; Koehler, 1998), some form of pre-publication approval, however lax, almost always exists for all along the continuum. We thus seek here a better solution, designed especially for internet technology firms, with highly valuable, highly informative, and highly sensitive data.

There are certainly other models of industry-academic relations (Perkman et al., 2013; Ankrah and AL-Tabbaa, 2015). For example, social scientists now often leave the academy to conduct research in industry or work in collaboration with data scientists within firms. These researchers then have far more access to the data and more influence on products that affect millions of people, but they must work on projects that are of the highest priority to the company. As a result, their work is not always aligned with the questions that are of greatest interest to the scholarly community.

The dilemma is nevertheless stark: Academic independence with inadequate information generates little value. Full information with dependence on the firm for pre-publication approval poses actual or apparent conflicts of interest. And the many ad hoc approaches around these problems are difficult for individual scholars to negotiate, especially on highly charged politicized issues. These issues are sometimes overcome with ad hoc approaches, but only rarely on issues that are politically or commercially sensitive. The challenge with all of the models described here is that they do not scale as well as they might. Data remains only accessible to relatively few, and scientific progress suffers as a result.

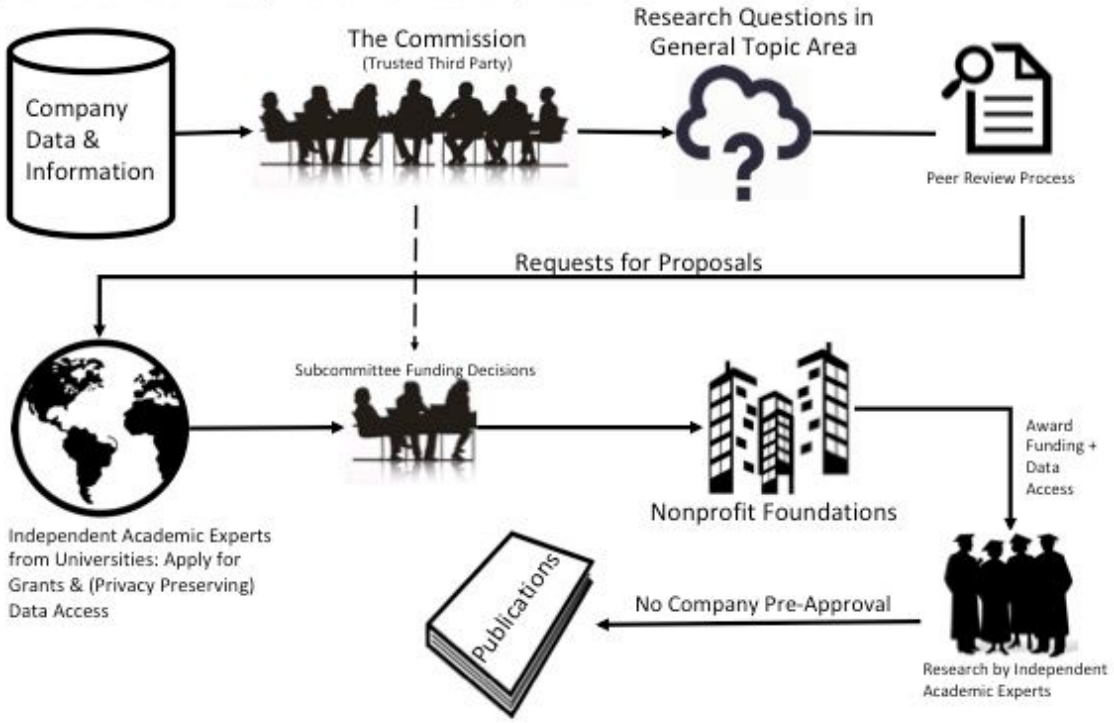
3. A Proposed Two-Part Solution

We now describe a plan designed to reduce the difficulties described in Section 2. Our model ensures that the company's interests are protected, the data is secure and kept private, and the researchers maintain independence. Our discussion in this section follows the outline in Figure 1.

The process begins by a company establishing a *general topic area* of research it is willing to pursue, such as the impact of social media on elections and democracy. Then within this area, we introduce a two-part organizational structure. In Part 1, we establish a *commission* composed of respected scholars which, as a trusted third party, will receive access to all relevant parts of the company's operations under confidentiality agreements. The commission will receive answers to all relevant questions about any platform, product, policy, or practice, including information

about the company's systems and data, that can help it achieve its goals. It will run relevant statistical analyses by working closely with the firm's data scientists who, in any organization, typically have a wealth of information not available in formal written documentation.

Figure 1: Outline of Industry-Academic Partnership Model



For both legal and privacy reasons, not all of the information shared with the commission will be made public. This is the innovation underlying the two-part structure: the commission will have access to all information required to make informed recommendations, and will filter this information to the broader research community on its own accord, omitting proprietary and other specifically delineated information follow agreed upon rules.

The process will only work if the commission has the trust of the broader academic community and general public. It is therefore designed to be composed of well-known, highly-respected, distinguished, senior scholars who represent the scientific community across important dimensions of demographic, political, substantive, and methodological diversity. At its discretion, the commission may also recruit experts in specific areas as academic consultants or for other specialized committees, such as about privacy or specific geographic regions. The commission and academic consultants will be compensated at fair market rates, and, in highly charged partisan or otherwise sensitive environments, paid by nonprofit foundations independent of the firm.

Because members of the commission are treated as insiders and given full access to sensitive information, they are not free to publish research without pre-publication approval, probably will not publish from this experience at all, and are explicitly prohibited from responding to requests for proposals or receiving funding as described below (which is another reason why only senior scholars should be recruited to participate on the commission). However, the commission has the ability and indeed obligation -- without permission from the firm -- to report to the public about whether the firm is keeping its end of the bargain, providing the commission full access, and answering all relevant questions. To be specific, if the commision concludes that the company has violated its agreement and prevented it from providing any piece of information it needs to address the general topic area, it will report this in a visible public statement. Once the commision is established, it will have a responsibility to regularly report on its activities and the firm's to the public, including decision-making criteria guiding the research agenda, scholar selection, and overall progress.

In Part 2 of our structure, after the commission gets up to speed on all relevant internal data systems, policies, platforms, and practices at the firm, it will identify a long series of important research questions, each of which it believes may be answerable with access to a specific, and privacy-preserving subset of the firm's data and systems, or with an appropriately and ethically designed new data collection procedure such as a randomized experiment. If answers to these questions can be ascertained from research on the platform and there are no legal or other agreed upon barriers to the research, then the commission will follow standard academic procedures and announce an open grant competition for *independent academic experts* to receive funding to take on this work. This competition will include formal, public requests for proposals, peer review, and "revise and resubmit" processes. The commission will appoint a subcommittee, possibly including others, to participate in the peer review process, and the commission will award grants based on input from this process. The firm will have no role in choosing outside experts or making funding decisions. When grants are awarded, the independent academic experts will receive funding through standard university procedures for sponsored research and data access from the company. The process for academics will thus be familiar and simpler than most other industry-academic partnerships, as they will be simply applying for and receiving a grant from a nonprofit foundation.

Thus, academics may participate in the process we have set up in several ways. They may be selected to serve on the commission or as consultants, apply to be independent academic experts, or participate as peer reviewers. The commission will also solicit advice at several stages of the process and so academics may choose to provide help there as well.

4. Application to Facebook

We developed our proposed model of industry-academic partnerships in the context of building a partnership with Facebook in the highly charged partisan atmosphere surrounding the issue of foreign influence through social media in the 2016 US presidential elections and the UK Brexit referendum, immediately following the Cambridge Analytica scandal. Because of how sensitive these issues are, we helped arrange for the funds for grants to independent academic experts to come from a politically and substantively diverse coalition of nonprofit foundations. The coalition includes the John and Laura Arnold Foundation, the Democracy Fund, the William and Flora Hewlett Foundation, the John S. and James L. Knight Foundation, the Charles Koch Foundation, the Omidyar Network, and the Alfred P. Sloan Foundation. (Since large companies have considerably more money than nonprofit foundations, we may at some point have additional funds flow directly from the company to grantees, but regardless, the company will remain disconnected from all granting decisions to the independent academic experts.)

We have also arranged for the Social Science Research Council (SSRC) to oversee the peer review processes for our project. Working with SSRC, the commission will appoint a subcommittee, possibly including others, to participate in the peer review process, with the commission having ultimate decision making authority on awarding these grants.

By mutual agreement, the general topic area for our project will be responsible research on the implications of social media and digital technologies in the world -- starting with democracy and elections. Here is Facebook's description:

> "The focus will be entirely prospective, with the goals of understanding Facebook's impact on upcoming elections — including in Mexico, Brazil, India, and the United States — and informing future decisions. The research sponsored by this effort will have benefits both for our understanding of social media's effects on democracy and for Facebook to better understand whether it has the right systems in place, i.e. are we effectively able to fight the spread of misinformation and foreign interference? Specific topics may include misinformation; polarizing content; promoting freedom of expression and association, protecting domestic elections from foreign interference; and civic engagement."

5. Participation and Cooperation

The structure of any industry-academic partnership must satisfy the company's legal and fiduciary requirements while also ensuring that academics can exercise independence in addressing the core challenges within the general topic area.

The company's influence over the process thus includes (a) selecting the commission, in agreement with funders and members as they are appointed; (b) agreeing with the commission on what topics are feasible to study on the platform, given personnel and other resource constraints, with available, privacy-preserving data, and (c) retaining the ability to exercise control in the event a research project would violate the company's legal obligations, interfere with ongoing or imminent litigation, violate privacy, or compromise proprietary information. The commission will ensure that the definition of research questions and formal requests for proposals not be so narrowly stated that they also predetermine the answer. In that circumstance -- where no one is vulnerable to being proven wrong -- nothing of value to science can be learned, and academics would be uninterested in participating. At the same time, few companies would participate in facilitating research designed solely to evaluate its own actions, at least not without the ability to learn how to improve its products or services going forward. If either of these outcomes seems likely, no industry-academic partnership is possible. And if the company constrains questions in ways that violate this agreement, the commission will follow its obligation described above to report to the public that its broader goals will not be met because of actions taken by the company.

Academics would probably prefer that companies have fewer rights to choose questions, so they can access any data they wish, but no private company will (or could legally) go forward without these rights. Insisting that they give them up would mean no data-sharing without pre-publication approval at all, which gets us nowhere. Moreover, just as many requests for proposals from nonprofit foundations and governments allow for only a circumscribed set of topics they choose, companies have this right here as well. For-profit and not-for-profit organizations may have different motivations for providing funding for certain questions and not others, but requests for proposals from all are constraining to some degree.

The optimal way forward, then, is to find research questions that are of intellectual interest to the scientific community and either provide valuable knowledge to inform product, programmatic, and policy decisions, or are orthogonal to company interests. Of course, any company participating in this process must understand that the point of research is to learn new things,

discover answers to existing questions, and find new questions never before conceived. As a result, some bad news for the company will sometimes unavoidably surface.

6. Principles

Many specific procedures need to be set up under this structure but throughout we hope all who who follow this new model will adhere to five principles in addition to those described above. We will adhere to them in the partnership we have forged.

First, to submit proposals, all independent academic experts must go through their standard ethical review processes such as the Institutional Review Board procedure in the U.S. and parallel processes in other countries.

Second, the privacy of individuals represented in firm data must be protected. Any breach in this system would damage the credibility of the researcher, the process we have arranged to make data available for the academic community, the social good intended by the researcher, and the reputation of the company. As such, researchers' access to and use of such data will be held to higher standards for privacy, confidentiality, and security than required by any existing law or university practice. Likely, we will not distribute data to outside experts to use in their computer systems, but instead will have these experts access minimally necessary data on company infrastructure with specialized, locked-down systems with continuous auditing. We will have industry and academic experts on privacy and security involved from the outset. On-site usage is one approach, but more convenient procedures for researchers are now available to provide "virtual clean rooms" on single use, highly secure laptops. Privacy and security will be improved continuously as new technology emerges and the project proceeds.

Third, we will require all funded research to follow the "replication standard" and thereby produce and archive replication data files (King, 1995). This means that published research completed under grants from this process will be replicable by other researchers, under specialized conditions which we will develop and publish. The privacy and confidentiality concerns of this research obviously complicate this process, but we have several procedures available to us we will use. For example, a formal citation will be established for every data set with a "universal numeric fingerprint" that uniquely identifies a dataset even if the format in which it is stored changes (Altman and King, 2007) and a persistent identifier. The code and methodological details (but not data) must be publicly available in Dataverse (see dataverse.org). And the full replication archive, including the data and all procedures necessary to replicate the analysis will be available internally at the company, where we will arrange access for academics.

Fourth, the independence of academic researchers needs to be protected as much as possible. The implementation of our model with Facebook on issues about democracy and elections, in particular, benefits considerably from the extraordinary participation of a large number of high profile, politically and substantively diverse nonprofit foundations. Having their endorsement, guidance, and funding -- and just the fact that they are working together with singular purpose to make this project a success -- eases considerably the difficulty of forging a partnership in the highly politicized domain of democracy and elections. In other industry-academic partnerships -- such as with smaller companies, in less politicized environments, or in substantive areas where funding from nonprofit foundations is unavailable -- our plan's two part structure would still work, with few modifications and without any added difficulties. The only real difference would be that the company may fund the commission, consultants, and outside experts directly.

Finally, we wish to emphasize that in choosing members of our commission, consultants, peer reviewers, and independent academic experts, several dimensions of diversity will be essential. These include diversity in methodological approach, substantive area, geographic region, race, ethnicity, gender, ideology, and political party.

7. Concluding Remarks

One might reasonably wonder whether now is, in fact, the time to discuss a data sharing program between internet companies and academics. Concerns about privacy are rightly at the forefront of everyone's mind in the wake of recent revelations. After all, the most notorious recent scandal began with a breach by an academic (acting as a developer) of a developer's agreement with Facebook, which barred his sale of the data to a for-profit company. That scandal is an academic scandal as well.

Yet, for this reason, now is precisely the time to have this conversation and to set up structures that protect users' privacy while allowing independent academic analysis of social media data. If we do not set up these institutions, only the firms themselves will have access to the data on some of society's most pressing challenges, including, most immediately, the impact the new social media environment has on democracy. Absent such an effort, many of those outside of large technology companies, including academics, commentators, and government regulators, will continue to distrust the companies' representations that they understand the extent of the problems, have conveyed them accurately, or have implemented adequate solutions. With this new approach, however, we can take a critical step toward independent analyses of the dynamics of social media's effect on society, which will have downstream benefits for both the general public and the firms, and can begin to tackle numerous other societal problems.

The key features of our approach include independence of scientific research from undue private influence; access to a major company's data under conditions incentive-compatible for the firm, individuals, and nonprofit organizations; a model for industry-academic engagement tailor-made to the unusual nature of the firms and the data they house; and deployment of the scientific community to help advance societal good with previously inaccessible information. Achieving these difficult objectives requires, for any one implementation of our model of industry-academic partnerships, delicate and often extensive negotiations throughout the process of structural organization, question definition, empirical research, and eventual publication. But the questions are too important, the potential advances too large, the range of knowledge that could be learned too significant, and the information about the greatest challenges of society too valuable to miss the advances the scientific community can bring to the table.

References

Altman, Micah and Gary King. 2007. "A Proposed Standard for the Scholarly Citation of Quantitative Data." *D-Lib Magazine*, 13. http://j.mp/2ovSzoT

Ankrah, S. and Omar, A.T., 2015. "Universities–industry collaboration: A systematic review". *Scandinavian Journal of Management*, 31(3), pp.387-408.

Banaji, Mahzarin R., and Anthony G. Greenwald. 2013. *Blindspot: Hidden biases of good people.* New York, NY: Delacorte Press

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen. 2017. *The value of flexible work: Evidence from uber drivers*. No. w23296. National Bureau of Economic Research, http://bit.ly/2HjgqPq.

Corzo, Janet and Perkins Eastman. 4/20/2015. "How Academic Institutions Partner with Private Industry" *R&D*, http://bit.ly/IndAc.

Jasny, B. R., N. Wigginton, M. McNutt, T. Bubela, S. Buck, R. Cook-Deegan, T. Gardner et al. 2017. "Fostering reproducibility in industry-academia research." *Science* 357, no. 6353: 759-761, http://bit.ly/2Hezzlx.

King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics*, 28, Pp. 444-452, http://j.mp/2oSOXJL.

Koehler, J.J., 1993. "The influence of prior beliefs on scientific judgments of evidence quality". *Organizational behavior and human decision processes*, 56(1), pp.28-55.

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen. 2017. *The value of flexible work: Evidence from uber drivers*. No. w23296. National Bureau of Economic Research, http://bit.ly/2HjgqPq.

Perkmann, Markus, Valentina Tartari, Maureen McKelvey, Erkko Autio, Anders Broström, Pablo D'Este, Riccardo Fini et al. "Academic engagement and commercialisation: A review of the literature on university–industry relations." *Research policy* 42, no. 2 (2013): 423-442.

Wilson, Timothy D., and Nancy Brekke. 1994. "Mental contamination and mental correction: Unwanted influences on judgments and evaluations," *Psychological Bulletin* 116(1):117.