

Restructuring the Social Sciences: Reflections from Harvard's Institute for Quantitative Social Science

Gary King, *Harvard University*

ABSTRACT The social sciences are undergoing a dramatic transformation from studying problems to solving them; from making do with a small number of sparse data sets to analyzing increasing quantities of diverse, highly informative data; from isolated scholars toiling away on their own to larger scale, collaborative, interdisciplinary, lab-style research teams; and from a purely academic pursuit focused inward to having a major impact on public policy, commerce and industry, other academic fields, and some of the major problems that affect individuals and societies. In the midst of all this productive chaos, we have been building the Institute for Quantitative Social Science at Harvard, a new type of center intended to help foster and respond to these broader developments. We offer here some suggestions from our experiences for the increasing number of other universities that have begun to build similar institutions and for how we might work together to advance social science more generally.

The social sciences are in the midst of an historic change, with large parts moving from the humanities to the sciences in terms of research style, infrastructural needs, data availability, empirical methods, substantive understanding, and the ability to make swift and dramatic progress. The changes have consequences for everything social scientists do and all that we plan as members of university communities.

Universities, foundations, funding agencies, nonprofits, governments, and others have been building social science research infrastructure for many years and in many forms, but recently a growing number of research universities have been organizing their response to the new challenges with versions of a new type of institution we created at Harvard, the Institute for Quantitative Social Science (IQSS; see <http://iq.harvard.edu>). As representatives from many universities have contacted or visited us to learn about how we built IQSS, and an increasing number have started their own related centers, we offer here some thoughts on our experiences to help distribute the same information more widely.

Gary King is the Albert J. Weatherhead III University Professor, and Director, Institute for Quantitative Social Science, Harvard University. He can be reached at GaryKing.org; king@harvard.edu.

In the sections that follow, we offer a summary of the changes remaking the social sciences, a brief overview of IQSS, and some suggestions for universities and their local academic entrepreneurs attempting to improve their social science infrastructure. Ultimately, universities build locally and cooperate internationally; as a result, the social sciences, each of the disciplines within it, what we all learn, and our impact on the world are all greatly improved as a result.

THE STATE OF SOCIAL SCIENCE

Recent Progress

The influence of quantitative social science (including the related technologies, methodologies, and data) on the world in the last decade has been unprecedented and is growing fast. Defined by the subset of “big data” (as it is now understood in the popular media) that has something to do with people, it is something every social scientist should feel proud to have contributed to. Indeed, few areas of university research approach the impact of quantitative social science. It had a part in remaking most Fortune 500 companies; establishing new industries; hugely increasing the expressive capacity of human beings; and reinventing medicine, friendship networks, political campaigns, public health, legal

analysis, policing, economics, sports, public policy, commerce, and program evaluation, among many others areas. The social sciences have amassed enough information, infrastructure, methods, and theories to be making important progress in understanding and even ameliorating some of the most important, but previously intractable, problems that affect human societies. Popular books and movies, such as *Moneyball*, *SuperCrunchers*, and *The Numerati*, have even gotten the word out.

An important driver of the change sweeping the field is the enormous quantities of highly informative data inundating almost every area we study. In the last half-century, the information base of social science research has primarily come from three sources: survey research, end-of-period government statistics, and one-off studies of particular people, places, or events. In the next half-century, these sources will still be used and improved, but the number and diversity of other sources of information are increasing exponentially and are already many orders of magnitude more informative than ever before. However, big data is not only about the data; what made it all possible are the remarkable concomitant advances in the methods of extracting information from, and creating, preserving, and analyzing those data and the resulting theoretical and empirical understanding of how individuals, groups, and societies think and behave. See King (2009, 2011).

A promising side effect of this change in research style is that the most significant division within the social sciences, that between quantitative and qualitative researchers, is showing signs of breaking down.

Although the immediate and future consequences of these developments for the world seem monumental, our narrower focus here is on the important consequences of these changes for the day-to-day lives of the social science faculty and students who support these efforts, and for the universities and centers that facilitate them. Social scientists are now transitioning from working primarily on their own, alone in their offices—a style that dates back to when the offices were in monasteries—to working in highly collaborative, interdisciplinary, larger scale, lab-style research teams. The knowledge and skills necessary to access and use these new data sources and methods often do not exist within any one of the traditionally defined social science disciplines and are too complicated for any one scholar to accomplish alone. Through collaboration across fields, however, we can begin to address the interdisciplinary substantive knowledge needed, along with the engineering, computational, ethical, and informatics challenges before us.

Many examples of the types of research that improved social science infrastructure makes possible are given in King (2009, 2011), but consider three that have been conducted with IQSS infrastructure in recent years.

First, for almost a century scholars have been studying what newspaper advertisements convey about social attitudes, purchasing patterns, and economic history (Salmon 1923). Until recently, the largest collection included a data set with only about 200 ads per year (Schultz 1992). Today, traditional newspapers, now operating online, display dynamic advertisements where ad content is

highly personalized. No two experiences on a newspaper website are likely to generate the same ad. With the resources available at IQSS, a faculty member archived more than 120,000 advertisements and documented how ad content changes as readers search for different first and last names. She found clear evidence of racial discrimination in ad delivery, with searches of names with a first name given primarily to black babies, such as Tyrone, Darnell, Ebony, and Latisha, generating ads suggestive of an arrest 75%–96% of the time. Names with first names given at birth primarily to whites, such as Geoffrey, Brett, Kristen, and Anne, generated more neutral copy: the word “arrest” appeared 0%–9% of the time, regardless of whether the actual subjects actually had an arrest record (Sweeney 2013).

Second, the quality of US state voter registration lists, from which the eligibility of voters is determined, has long been an issue in American politics. Yet, the data requirements meant that previous systematic analyses of this have been one-off studies of small numbers of people or places. More recently, two faculty members and a team of five graduate students from IQSS tackled this problem by studying all 187 million registered voters from every US state (Ansolabehere et al. 2013). They found that *one third* of those listed by states as “inactive” actually cast ballots, and the problem is not politically neutral. The researchers have gone on to suggest productive solutions to the problem.

And finally, fewer than two decades ago, Verba, Scholzman, and Brady (1995) amassed the most extensive data set to date on the voices of political activists, including 15,000 screener questions and 2,500 detailed personal interviews, and they wrote a landmark book on the subject. Shift forward in time and, with new data collection procedures, statistical methods, and changes in the world, an IQSS team composed of a graduate student, a faculty member, and eight undergraduate research assistants were able to download, understand, and analyze all English language blog posts by political activists during the 2008 presidential election and develop methods capable of extracting the meaning from them (Hopkins and King, 2010). The methods were patented by the university and licensed to a startup, and now mid-sized, company (Crimson Hexagon, Inc.). Even more recently, a team of two graduate students, a faculty member, and five undergraduates downloaded 11 million social media posts from China before the Chinese government was able to read and censor (i.e., remove from the Internet) a subset; they then went back to each post (from thousands of computers all over the world, including inside China) to check at each time point whether it was censored. Contrary to prior understandings, they found that criticisms of the Chinese government were not censored but attempts at collective action, whether for or against the government, were censored (King, Pan, and Roberts 2013).

In these and many other projects, IQSS scholars built methods and procedures that made it feasible to understand much larger quantities of information than could possibly have been accessed

by earlier researchers. These research projects depended on IQSS infrastructure, including access to experts in statistics, the social sciences, engineering, computer science, and American and Chinese area studies. Having this extensive infrastructure and expertise frees researchers associated with IQSS, and affiliated scholars, to think more expansively and to take on projects that would not merely have been impossible otherwise, but which we would have likely not even imagined.

The Coming End of the Quantitative-Qualitative Divide

A promising side effect of this change in research style is that the most significant division within the social sciences, that between quantitative and qualitative researchers, is showing signs of breaking down. You can almost hear the quantitative researchers—who have spent decades analyzing time series cross-national data sets with only a few impoverished variables—saying “OK, we give! So much is left out of our models that qualitative researchers include. Can’t someone systematize that information so we can include it?” And at the same time, you can just about hear the qualitative researchers complaining “We are overwhelmed by all the information we are gathering, and more is coming in every day; we can’t read, much less understand, more than a tiny fraction of it all. Can’t you quantitative researchers do something to help?” In fact, versions of both are commonplace within the context of numerous individual research projects.

Fortunately, social scientists from both traditions are working together more often than ever before, because many of the new data sources meaningfully represent the focus and interests of both groups. The information collected by qualitative researchers, in the form of large quantities of field notes, video, audio, unstructured text, and many other sources, is now being recognized as valuable and actionable data sources for which new quantitative approaches are being developed and can be applied. At the same time, quantitative researchers are realizing that their approaches can be viewed or adapted to assist, rather than replace, the deep knowledge of qualitative researchers, and they are taking up the challenge of adding value to these additional richer data types.

The divergent interests of the two camps also converge as the need for tools to cope with, organize, preserve, and share this onslaught of data, the search for new understandings of where meaning exists in the world and how it can be represented systematically, and the rise of inherently collaborative projects where researchers bring their own knowledge and skills to attack common goals. Instead of quantitative researchers trying to build fully automated methods and qualitative researchers trying to make do with traditional human-only methods, now both are heading toward using or developing computer-assisted methods that empower both groups. This development has the potential to end the divide, to get us working together to solve common problems, and to greatly strengthen the research output of social science as a whole.

The Boundaries of “Social Science”

As social science has become increasingly interdisciplinary and collaborative, so too has the de facto definition of the field broadened. The result is that the historical or institutional definitions of “social science,” based only on what work is being done in specific departments (sociology, economics, political science, anthropology, psychology, and sometimes others), is unhelpful as it excludes numerous social scientists elsewhere in most uni-

versities. We instead use the term “social science” more generally to refer to areas of scholarship dedicated to understanding, or improving the well-being of, human populations, using data at the level of (or informative about) individual people or groups of people.

This definition covers the traditional social science departments in faculties of schools of arts and science, but it also includes most research conducted at schools of public policy, business, and education. Social science is referred to by other names in other areas but the definition is wider than use of the term. It includes what law school faculty call “empirical research,” and many aspects of research in other areas, such as health policy at schools of medicine. It also includes research conducted by faculty in schools of public health, although they have different names for these activities, such as epidemiology, demography, and outcomes research.¹

The breadth of the field also covers many of those with whom we collaborate when they spread social science to their fields. To take one such example, over the last 20 years, political methodology has built a bridge to the discipline of statistics and the methodological subfields of the other social sciences (such as econometrics, sociological methodology, and psychometrics). Scholars, who began by importing methods from those fields, now also regularly make contributions used in those fields as well. Political science graduate students are now trained at a high enough level in political methodology so that they can move from the end of a sequence in political science directly into advanced courses in these other fields. Students in these other fields also do the same in our courses. The resulting vibrant interdisciplinary collaborations have resulted in statisticians and others becoming participants in the enterprise of social science.

Another version of the same pattern is now beginning to emerge between several traditional social science disciplines and computer science. Graduate students in economics, political science, and sociology now regularly learn computer languages and computer science concepts, and they are even beginning to include formal training in computer science as part of their graduate degrees. Associated with this development are computer scientists doing research in what is effectively social science. Indeed, this activity is being formalized in new departments at some universities, often under the banners “computational social science,” “applied computational science,” or “data science.”

Any scholar doing research in the area, regardless of their home department, should be included in a proper definition of social science. In fact, strictly speaking, parts of the biological sciences are effectively becoming social sciences, as genomics, proteomics, metabolomics, and brain imaging produce huge numbers of person-level variables, and researchers in these fields join social scientists in the hunt for measures and causes of behavioral phenotypes. These fields developed very differently from the social sciences, but they now use many of the same survey instruments, statistical methods, substantive questions, and even data sets. When methods, data, procedures, theories, and institutions can help research in other areas, the more inclusive we are and the more we will all benefit.

WHAT TYPE OF CENTERS TO BUILD

In this section, we describe the key elements behind IQSS and related social science research centers. We describe how community is the fundamental driver behind successful centers; how to

build such a community even though individual faculty members may well be pursuing their own divergent self-interests most of the time; and the standard elements of successful centers. We focus on how turning the insights of social science research on ourselves can greatly increase the chances of success. Ultimately, social science centers, run by social scientists who are familiar with the social science literature, have tremendous advantages not usually available to those building other types of university centers.

The Goal

We began IQSS with a research project, asking a wide range of academic leaders what distinguished the world's most renowned academic research centers—in their heyday, the Cavendish, Bell Labs, ISR, some of the Population Centers, and so on—from others, and what was the key ingredient for their success. Most said more or less the same thing in different ways: yes, you need the obvious components such as space, money, staff, colleagues, and projects, and of course the end product in terms of the creation, preservation, and distribution of knowledge is the ultimate measure of success. However, by common assent (although often in very different languages), by far the most important component leading to success identified was *community*. The world's best research centers each had an enviable research community that caused individual scientists to want to join in and contribute. Members of the community joined for either the self-centered

research. At the other extreme, IQSS provides what is sometimes thought of as “mere” administrative or infrastructural services, such as grant support that enables scholars to focus only on the intellectual component of proposals (leaving the rest to our expert staff); help fixing computer code, desktop computers, cell phones, or survey questions; or assistance incubating, administering, and hosting centers, labs, research groups, student and scholarly activities, and technology platforms. Although the former extreme may sometimes be more fun than the latter, activities all along this continuum are valuable. They all attract scholars to IQSS who might not otherwise have come, leading to synergies we would not otherwise have been able to realize. Plumbing may not be the most intellectually stimulating activity, but if the sewage pipes in your house break, the plumber becomes the most important person around. We are proud to provide “plumbing services” right next to someone who can help you prove the theoretical properties of a new statistical estimator, because they will each get you to visit IQSS, to interact with others, and to give back.

We therefore aim in the first instance to help individual faculty and students get their work done better, faster, and more efficiently on their own terms. Then, while individual scholars are receiving these individual services for their narrow self-interests, they cross paths with other researchers often from apparently distant areas, find collaborative opportunities, and eventually make

However, by common assent (although often in very different languages), by far the most important component leading to success identified was community. The world's best research centers each had an enviable research community that caused individual scientists to want to join in and contribute.

reasons of maximizing the quality of their own research, or because (as social science teaches) social connections provide independent motivation. Either way, the quality of the community is fundamental.

Adjusting Individual Incentives to Build Community

How do we create a community out of large numbers of ambitious, hard driving, often single-minded, researchers pursuing their own separate, and often competitive, research goals? Our answer, and our operating theory, is, at the first instant, to make IQSS attractive to individual researchers by ensuring that the specific services, products, and programs they can access make their research better and the research process faster and more efficient. Faculty and students often *come* to IQSS as individuals to solve their own highly specific problems holding back their research; they then *stay* for the research community. The advantages of the community then feeds back on itself, improving IQSS for those already participating, and then providing independent motivation for them to stay and others to join.

The services, products, and programs that IQSS offers researchers fall on a continuum from academic to administrative. At one extreme, we developed a convening power that attracts some of the world's best social scientists from Harvard and elsewhere to spend time here and interact at a very high level about their

substantial contributions to building our research community. Every path that is crossed increases the probability of an intellectual connection, even if each path had nothing to do with the reason for the connection. Individual scholars are not always focused on, or even aware of, their important contribution to the collective, but the research community is much stronger as a consequence of these interactions. The result is that the community here, and in similar organizations, seems to be flourishing and is now filled with social scientists from disciplines representing the many departments and schools at Harvard and beyond.

Organizing the Institute

We organize IQSS activities into what we offer scholars: *research programs, services, and products*. Our research programs include the Program on Quantitative Methods, Program on Survey Research, Program on Text Research, Experience Based Learning in the Social Sciences, the Data Privacy Lab, undergraduate and graduate scholars programs, the NASA Tournament Lab, the Global History of Elections Program, among others. Larger entities under the IQSS umbrella also include the Center for Geographic Analysis, the Murray Research Archive, and the Harvard-MIT Data Center.² These centers and programs offer numerous weekly seminars, regular workshops, and one-off or recurring research conferences. Hundreds of people come and go on a regular basis.

Services involve common administrative management for all the separate research groups, such as financial management and transaction approvals, strategic advice, human resources, and technology infrastructure, including support for acquiring, storing, and analyzing data on our high-performance computing cluster; research technology consulting, technical training classes, desktop support, public labs, and the like. We also prepare pre- and post-award sponsored research administration, following the theory that the only part of grants faculty should have to write is the intellectual justification. The key to gaining the considerable economies of scale possible from this activity is pairing common administration and management with intellectual leadership left entirely in the hands of separate faculty leaders in charge of each program. The faculty members get to focus on what they are good at and benefit from staff focusing on what they are good at. And all the while the institute benefits by the economic efficiencies gained and the community that is fostered.

Our products include services that we packaged and made self-service. These include the Dataverse Network, OpenScholar, Zelig, a research computing environment, and others, some of which we discuss below.

Applying Social Science Research Findings to Ourselves

To best facilitate the types of researcher connections that foster community, we founded IQSS as an unusual *hybrid organization*, both a research center and an integral part of the university admin-

To best facilitate the types of researcher connections that foster community, we founded IQSS as an unusual hybrid organization, both a research center and an integral part of the university administration. We often do both together by taking routine activities of the administration and turning those into quasi-research projects.

istration. We often do both together by taking routine activities of the administration and turning those into quasi-research projects. Good social science research centers are not merely generic research centers, functionally equivalent to those in other fields save only for the subject area. The fact that we are behavioral scientists gives us an inherent advantage in understanding, building, and running organizations, in designing policies that build off individual incentives, and in fostering intellectual communities. And the fact that we have technical computer and statistical skills means we also have an advantage in automating routine tasks. Together these advantages extend the impact, efficiency, creativity, and productivity of the overall effort.

For example, by applying quantitative social science research techniques and cutting edge computer science to our own activities, we can sometimes make products that scale to many more faculty members and students at far lower cost—improving the research lives of those associated with IQSS and freeing up funding for “higher level” research activities. For example, we automated, through our Dataverse Network® software project (see <http://TheData.org>; Crosas, 2011, 2013; King, 2007), most of the activities of the Murray Research Center (previously at the Radcliffe Institute and now at IQSS). For more than three decades, the Murray was widely known for carefully and lovingly collect-

ing and curating a small group of data sets. By automating the operations of the Murray, the staff became far more efficient.

In addition, the Murray’s previously traditional model of data collection was similar to many other archives, but not well aligned with the incentives of researchers. Researchers who wanted to make data available had to choose between putting it in a professional archive like the Murray—which ensured long-term preservation, but often resulted in citations thanking the Murray rather than the researcher—or distributing it themselves—which would keep credit with the researcher but would likely flout professional archiving standards and so usually give up long-term preservation. The Dataverse Network project breaks this tension by using better technology and aligning it properly with incentives gleaned from social science research: we do this by adding an extra page to any researcher’s website with a virtual archive, called a “dataverse.” The dataverse includes a list of the researcher’s data sets, along with a vast array of services, including archiving, distribution, on-line analysis, citation, preservation, backups, disaster recovery, among others. The researcher’s dataverse page devolves all credit to the researcher by being branded entirely as the researcher’s (with the look and feel of the rest of the researcher’s website) but the page is virtual and so installation takes a few minutes, and it is served out by a central archive and managed by others following professional archiving standards. We also researched citation standards and developed a standard for data, so that the researcher who makes data avail-

able through dataverse gets more web visibility and more academic credit (Altman and King, 2007).

In the first year after the Murray moved to IQSS, it collected more than 10 times the number of data sets as had been collected in the previous 30 years at Radcliffe, at lower cost, and with vastly increased access to data for our researchers and others. At the same time, we directed some of the archive’s financial resources to more productive research activities. The synergies from this activity are apparent in the ecosystem of research projects from around the world that have grown up around dataverse, the many scholars who contribute to and work collaboratively with this open source software project instead of building their own solutions from scratch, and the millions of dollars in federal and other grants that have supported these activities. The Dataverse Network now offers access to more social science research data than any other system in the world. The Harvard University library system has also formally adopted the Dataverse Network and is using it to provide archiving services to astronomers, biologists, medical researchers, humanists, and others. The open source software is also installed at a variety of other universities around the world.

We have also repeated this model several times in other areas. In each, we find a piece of the administration, a center, or an

activity, and we apply social science methods, theories, technologies, evaluation procedures, and insights about human behavior to improve the resulting services or activities.

Because we are emphasizing the advantages of “plumbing,” consider an example near this end of the continuum—desktop computer support, an essential but thankless activity, typically engendering many complaints, flames, and turmoil. We fixed these problems by setting up a system that encouraged the staff to “teach to the test.” To be more specific, we customized (through considerable experimentation) an automated ticketing system, and tuned the incentives with what we know from social science research. Thus, when faculty, students, or staff have a desktop support issue, they e-mail the support group and receive an automated response immediately and a promise of a human contact shortly. If a member of the team does not make contact within that time period, they get prompted automatically. Because their manager would get prompted too, users rarely have to wait long. If during the interaction, the staff member is waiting for information from the user or the user is waiting for something from the staff member, the ticketing system gently prompts the right person to make sure progress is made. When the staff member thinks the issue has been resolved, the system administers a fast three-question “how did we do” survey. If a user does not mark “extremely satisfied” for all three questions, then staff, their supervisors, and top management are notified immediately. Staff closely monitor how they do on these brief surveys and try to satisfy users as indicated by the questions; by constantly evaluating and tweaking the survey questions, the staff, management, and users understand each other much better. And, after some years of learning, and randomized experiments, it now seems to work well. In the last 18 months, the number of tickets marked “dissatisfied” or “extremely dissatisfied” (of more than 7,000 filed) is exactly zero. Users are never left wondering what is happening, and staff know exactly what the community regards as good service. With the management of desktop support thus effectively automated, the rest of us can turn from firing off angry memos about customer service to writing more scholarly articles.

When possible, we emphasize *infrastructure that scales*, so that spending is highly leveraged. We do this by our focus on research computing infrastructure that is naturally amenable to use by large numbers; by our day-to-day emphasis on creating synergies among the different parts of the institute; with the help from faculty and students from all over the university who interact here; and by marshaling the efforts of several open source communities in contributing software and other assistance from inside and outside of Harvard. Other examples of these activities include OpenScholar (<http://openscholar.harvard.edu>), a single open source (software as a service) web software installation that creates thousands of highly professional and customized websites for faculty, projects, and academic departments, saving \$6000–\$25,000 per site (as of this writing, about 3,000 scholars and departments have OpenScholar sites at Harvard, and about 150 other universities have their own installations); “Zelig: Everyone’s Statistical Software” (<http://projects.iq.harvard.edu/zelig>), an all-purpose statistical package built on the R Project for Statistical Computing, now used by hundreds of thousands of researchers worldwide; our “Research Computing Environment,” which is an infrastructure to make high-performance research computing straightforward to run, and easier to scale; among others.

Other Models

Centers elsewhere may choose to work on software infrastructure, like IQSS, and if so can work collaboratively with us on these projects, as some do now. As the social sciences branch out, get connected to other fields, and draw in new forms of data, they need many different types of infrastructure. Any of these approaches will likely benefit by applying social science principles and research to our own activities in these and many other ways.

SUGGESTIONS FOR ACADEMIC ENTREPRENEURS

Don’t Try to Replicate the Sciences

As parts of the social sciences move from the humanities to the sciences, we might wish to receive the level of support from our universities that our colleagues do in the natural and physical sciences. Social science research would certainly be massively better off if we outfitted all social science faculty members with their own lab on the scale of those in, say, chemistry or biology, with \$2–3 million of startup money, 3,500 square feet of lab space, and a dozen full-time employees. This is, of course, wildly unrealistic in the short term (and insisting your university administration instantly impose this notion of equity would likely get your more reasonable requests ignored), but we ought to be able to make such an expectation unnecessary as well. That is, instead of attempting to replicate the physical and natural science model within the social sciences today, we can take a far more efficient approach that involves building *common infrastructure* to solve problems across the labs and research programs. The fast emerging models of collaboration and cooperation make this both possible and much more likely to be productive.

Don’t Try to Build it from Scratch

Handing a copy of the IQSS budget to your university’s administration as your budget request to start your own center is highly unlikely to work. The dollar amounts are just too big for them to take you seriously, or for your administration to come up with the money to pay for it even if they want to. If we had sent what is now our budget as a proposal to the Harvard administration to form IQSS, they would have thought we were crazy, politically naive, or both.³ The point, however, is that we did not build IQSS from scratch; we built it from components that existed—largely unconnected—around the university. In our case, these included the Murray Research Center, the Harvard-MIT Data Center, the Center for Geographic Analysis, and some others. In most universities, a good deal is spent on social science infrastructure, but the parts are scattered under different administrative units, not working together, without any faculty direction, and each working less efficiently than they could together. Look for such units in the obvious places, but do not overlook the library, the information technology infrastructure, academic computing groups, and elsewhere in your administration.

A good approach is to carefully map out the local political landscape and find existing units that already have some type of financial support. Then talk to those individuals who are in charge of each unit and find out what they need, how to empower them, how they can accomplish *their* goals by working together with you. Radical decentralization is often the best politically achievable path to centralization. Build from the ground up and the specific request to the administration can be more reasonable and

easier to accomplish; instead of something they cannot approve, you can make it almost impossible to turn down.

Build Adaptable Infrastructure

The infrastructure we need in the social sciences must be reliable and flexible. Our field, and the technologies we use, is changing fast, as are, therefore, our infrastructural needs and research opportunities. For example, as technology changes, we adapt IQSS along with it. We change the organization charts regularly. In only the last few years, we have built large open source computer programs, started new seminar series, run international conferences, brought together scholars from disciplines who have rarely collaborated, taken over and built quasi-research projects that make our university administration more efficient, started new programs, closed down completed programs and projects, spawned commercial firms and nonprofit startups, given services we developed to other parts of the university to run, educated students and faculty in new technologies, data, methods, and theories, and led many other activities. We build, but we also continually rebuild.

Build Incentive Compatible Administration at Scale

Research institutes like IQSS, and its various component centers and programs, require substantial faculty time and effort. Faculty members may love to teach, but running the university, and espe-

economies of scale will eventually turn into diseconomies of scale, but as long as the staff is properly hired, managed, and organized, few social science centers are near that point. Managing a large staff may require different skills than managing only two, but with a proper hierarchical organization, the task need not be more difficult or time consuming.

Emphasize Extreme Cooperation

Tremendous progress can be made merely by cooperating with other units. This does not mean acquiescing to every request from the outside, because most other units will not make requests and collaboration needs to be incentive compatible on both sides. Instead, find other units and do whatever it takes to establish connections, collaborations, and joint activities that make sense. If academic research became part of the X-games, our competitive event would be “extreme cooperation”; administrative units within universities do best when they follow that lead, especially because so few do. A key to remember is that influence is more important than control. If you give up the idea of being the sole supplier and producer of every activity, you can have far more influence intellectually, educationally, socially, and politically. It is also generally worth cooperating for its own sake in the short run, even if it for a while it takes considerably more effort than the benefits received.

Faculty members may love to teach, but running the university, and especially research are also an essential part of their mission. So the only way to build infrastructure sustainably is to make it incentive compatible.

cially research are also an essential part of their mission. So the only way to build infrastructure sustainably is to make it incentive compatible. Buying off faculty with time off or extra compensation can work, but is not efficient and probably not sustainable. A better approach is to align the public spirited interests of the center with the private interests of the faculty leadership.

I have always been closely involved in computer operations because I teach methods and need my students to have the best possible computer technology. Leaving computer technology to the university IT department does not work, no matter how qualified they are because their incentives are first to satisfy the 95% with vanilla services, whereas cutting-edge methods researchers are usually in the remaining 5%. But the same holds true for many other areas: university bureaucracies are appropriately designed for the many people they serve, whereas researchers are by definition at the cutting edge and therefore need more finely tuned or different services.

Faculty involved in administration are at first fearful (for their time, research careers, etc.) of hiring staff and building administrative structures, but the economies of scale are valuable and when done properly incentive compatible, too. I think I got this point the first time I noticed two of my staff members going out to lunch to solve a problem without me. At that time, we had only two staff; now we have 50–100 (depending on how you count), but the economies of scale continue way beyond where we are now. With more staff, you can hire better people, build career paths, and so hire even more qualified people, and so on. Undoubtedly,

You Don't Want Overhead from Grants

Early in the negotiations to create many centers is a often a discussion about whether it can be funded with overhead from federal grants. My advice is to not raise the issue and to turn it down if offered. The goal is to build durable infrastructure, not meet a payroll and have to fire people with every grant you happen to lose. The library and student health services also do not pay permanent staff from overhead on grants they bring in. A much better setup is for the administration to make whatever commitment they desire. If you do bring in a lot of money in grants, some level of trust will mean that you can count on them being somewhat more generous the next time you have a request. This need not be set down in writing, or even said, but it will happen. As congressional scholars have discovered, it is better to shoot for favor not favors.

Keep a Role for Theorists

Because most of the advances in the social sciences have been based on improvements in empirical data and methods of data analysis, some argue that the theorists (economic theorists, formal theorists, statistical theorists, philosophers, etc.) have no part in this type of center. This makes no sense. In every social science field, and most academic fields, a friendly division exists between theorists and empiricists. They compete with each other for faculty positions and on many research issues, but all know that both are essential. The empiricists in your center will need to interact with theorists at some point, and the theorists will benefit by

conditioning their theories on better empirical evidence. The fact that the big data revolution has enabled more progress on the empirical front does not reduce what theorists can contribute.⁴

CONCLUDING REMARKS: FEDERAL FUNDING PRIORITIES

The social sciences are undergoing a renaissance, and the infrastructure making it possible is growing, adapting, and greatly furthering our collective goals. As we all separately nurture and build this infrastructure within our own universities, we should not lose sight of a set of logical national and international goals that are even broader. Toward this end, we should cooperate and further build connections across centers within different universities. Then at the right time, we should set a collective goal to work together to change federal funding priorities. (And I'm not talking about the short-sighted recent change which effectively allows the National Science Foundation to fund any political science research except that about members of Congress or the public policies they write.)

Instead, we should think broader, and bigger. Most federal research funding comes from the \$31 billion National Institutes of Health (NIH) and \$7 billion National Science Foundation (NSF) budgets; in this, the social sciences are relegated to merely 4.4% of the smaller NSF budget. Although portions of NIH and other NSF programs contribute to the broader social science research enterprise, the disparity between these federal spending patterns and congressional priorities is enormous. Although members of Congress are clearly interested in many areas of health, science, and technology, they must be focused on issues their constituents want or they will lose their jobs. The issues that concern Americans the most have long been those directly addressed by social scientists, including the economic, political, cultural, and social well being of themselves, their communities, and the country. Of course, Washington is not in the business of funding researchers because they study interesting topics. Only when we can demonstrate that we can make a real difference will the funds flow. As our impact on solving problems becomes more and more obvious, changing federal priorities to more seriously fund social science research will be easier to see as in everyone's interest. At the right point, we should all consider a road trip to Washington.

ACKNOWLEDGMENTS

Thanks to Neal Beck, Kate Chen, Mercè Crosas, Phil Durbin, and Mitch Duneier for many helpful comments and suggestions. ■

NOTES

1. Among public health scholars, the term social science is sometimes confused with, and so must be carefully distinguished from "social epidemiology," which is one of many subfields of our broader definition of the social sciences
2. We also included the now defunct Center for Basic Research in the Social Sciences.
3. By all means use whatever success we have had as evidence that your university needs to invest more to compete. And since the rules of our industry are symmetric, we hope you succeed!
4. Moreover, theorists don't cost anything! They require some seminars, maybe a pencil and pad, and some computer assistance. There is no reason to exclude them, and every intellectual (and political) reason to include them.

REFERENCES

- Ansolabehere, Stephen, Adam Cox, James Snyder, Anthony Fowler, Marshall Miller, Jordan Rasmusson, Benjamin Schneer. 2013. "Inactive and Dropped Records on State Voter Registration Lists." working paper.
- Altman, Micah, and Gary King. 2007. "A Proposed Standard for the Scholarly Citation of Quantitative Data." *D-Lib Magazine* 13 (3/4) <http://j.mp/ikyBfV>.
- Crosas, Mercè. 2013. "A Data Sharing Story." *Journal of eScience Librarianship* 1 (3): 173–79. <http://j.mp/12yqPez>.
- Crosas, Mercè. 2011. "The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data." *D-Lib Magazine* 17 (1–2). <http://j.mp/12yqVCZ>.
- Hopkins, Daniel, and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54 (1): 229–47. <http://j.mp/jNFDgI>.
- King, Gary. 2007. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." *Sociological Methods and Research* 36 (2): 173–99. <http://j.mp/iHJcAa>.
- King, Gary. 2011. "Ensuring the Data Rich Future of the Social Sciences." *Science* 331 (11): 719–21. <http://j.mp/mw64M8>.
- King, Gary. 2009. "The Changing Evidence Base of Social Science Research." In *The Future of Political Science: 100 Perspectives*, eds. Gary King, Kay Scholzman, and Norman Nie. 91–93. New York: Routledge. <http://j.mp/k5lI9s>.
- King, Gary, Jennifer Pan, and Molly Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107 (2): 1–18. <http://j.mp/LdVXqN>.
- Salmon, Lucy. 1923. *The Newspaper and the Historian*. New York: Oxford University Press.
- Schultz, M. 1992. "Occupational Pursuits of Free American Women: An Analysis of Newspaper Ads, 1800–1849." *Sociological Forum* 7 (4): 587–607.
- Sweeney, Latanya. 2013. "A Closer Examination of Racial Discrimination in Online Ad Delivery." Working paper. Data archived at <http://foreverdata.org/onlineads>.
- Verba, Sidney, Kay Scholzman, and Henry Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.