SYMPOSIUM

Technology, Data, and Politics

Editors' Introduction to Technology, Data, and Politics

Sheena Chestnut Greitens, *Harvard University, guest editor*
Gabriel Koehler-Derrick, *United States Military Academy, West Point, guest editor*

This symposium addresses the rise of Internet data—and, more broadly, the data that is generated by a range of information and communication technologies (ICTs). Global ICT growth could have wide-ranging implications for the study of politics, but political science is currently lagging behind other disciplines in its use of ICT-derived data—particularly compared to fields such as public health and economics (Nadav et al. 2011; Carneiro and Mylonakis 2009; Choi and Varian 2009; Christakis and Fowler 2009; Ginsberg et al. 2009; Jensen 2011). The articles in this symposium, therefore, focus on the question of how political scientists can and should think about using ICT-derived data in their research.

The lag in political science is all the more surprising given that the rise of information and communication technologies, coupled with a growing interest in "Big Data," has garnered significant attention in recent years from media, policy makers, and academics. Developments around the globe have inspired wide-ranging debates about the role of these technologies in political events, whether in the form of speculation (now largely debunked) about Twitter catalyzing political mobilization during the "Arab Spring," or recent claims that president Barack Obama's campaign used sophisticated online marketing and data analysis to identify and galvanize voters whose support led to his reelection. By providing large new data caches, information and communication technologies have also contributed to the rise of the "Big Data" phenomenon—exponential growth in new sources of digital data that researchers can exploit.

The challenge thus far, however, has been how to make use of this data. The pace of data creation has threatened to overwhelm researchers' abilities to capture and analyze it; *The Wall Street Journal* described the problem as one of "data obesity" rather than data scarcity (Finn 2012). There has also been, as yet, relatively little discussion of how these data and approaches to using them can be incorporated alongside traditional data and methodologies within a multimethod research design framework.

The contributions to this symposium, therefore, take the next step in transforming this widespread intellectual curiosity about the role of Internet- and ICT-derived data into a better-developed and more fruitful research agenda. They do so, at the most fundamental level, by introducing major new sources of data derived from a range of information and communication technologies, and by demonstrating new methods of analysis that can be brought to bear on these data—whether on their own or in combination with more familiar data sources and methods. The authors address concerns that span the traditional subfields of political science, considering questions in American politics, comparative politics, international relations, political theory, and methodology.

The articles simultaneously introduce several larger points about the nature of the emerging data in the field. They analyze how different political contexts affect the type of data that is produced and made available to researchers. They outline some of the variations present in ICT data and probe how these variations do or do not correlate with variations in offline behavior. They also identify scope conditions, highlighting where these new data and tools can be most useful as well as identifying their limitations. Collectively, they demonstrate how these data and analytical techniques can be combined with classical methodological tools and creative research design and incorporated into the broader toolkit available to political scientists.

ICT-DERIVED DATA: SOURCES AND METHODS

What kind of ICT-derived data exists, and how can political scientists use these different types of data? The articles in the pages that follow deal with a range of ICT-generated data, drawn from sources that will interest readers working across subfields and topics. Half of the contributions to this symposium—those by Granka, Koehler-Derrick, and Mellon—focus on Internet search data. The others go beyond search data: Hassanpour discusses Google's Ngram tool, a huge corpus of digitized books, Goldstein explores "Social fMRI" data generated by smartphones, and Greitens examines states' ICT policies themselves as a source of data.

doi:10.1017/S1049096513000255
https://doi.org/10.1017/S1049096513000255 Published online by Cambridge University Press

This list is far from exhaustive. Other research has examined sources including Facebook posts (Bond et al. 2012); Twitter feeds (Mostak 2012); cell phone coverage (Shapiro and Weidmann 2012); and non-Western blogs, microblogs, and website content (Bamman, O'Connor, and Smith 2012; King, Pan, and Roberts 2012; Nielsen 2012). We are confident that this symposium is only the first step in finding creative ways of using ICT technology and other new forms of electronic data to answer challenging research questions.

The contributions included here, however, do showcase some of the major advantages of that data: its consistency of format, the speed and ease with which it can be collected, and its high volume. The authors demonstrate that information and communication technologies can provide political scientists with vast quantities of microempirical data to examine changes in behavior over time, geographic relationships, and process questions about time and scale.

Much of the recent public attention to technology's importance in politics has been driven by the supposition that technology has a uniform and prodemocratic impact. Actually, however, it is characterizing and explaining variations *in the role of technology in political life that will sustain a robust research agenda.*

The articles that follow also use a range of analytical methods, some innovative and some more traditional, to leverage these data and gain new insights into political phenomena of interest. Koehler-Derrick uses Internet search data to analyze framing battles between Egyptian protestors and the military during a period of tense confrontations in late 2011, whereas Hassanpour relies on the NGram tool to analyze trends in political discourse. Granka uses search data to build and test a model for forecasting the American presidential elections and Republican primaries, while Mellon pairs search results with traditional survey data to discuss the challenges of validation. Goldstein explores the potential for Social fMRI data to be used in experimental projects on improving governance and development outcomes, and Greitens demonstrates how, through careful contextual study, ICT data can shed light on political developments inside even the most restrictive of authoritarian regimes.

ICT-DERIVED DATA: THEMES AND QUESTIONS

As political scientists begin to incorporate ICT-derived data into their work, what issues and potential pitfalls should they keep in mind? As we compared the symposium contributors' data sources and analytical approaches, several themes emerged. These emphasized both the potential that technological developments hold for political science research, and their current limitations. Throughout this symposium, therefore, we have sought to be clear about the strengths and weaknesses of particular data sources for different types of research inquiry, and to candidly discuss the best analytical frameworks for bringing that data to bear on questions of interest to political scientists.

The first theme that emerged was the need, when selecting methods of analysis and drawing conclusions, to pay attention to the political context within which ICT data are generated. In some of the symposium's articles, such as the contributions by Goldstein and Greitens, political factors clearly condition the adoption of technology and therefore may restrict or bias the generation of data from that country. Other contributions, such as the one by Mellon, stress the need to take linguistic and cultural differences into account.

The range of sources and methods used here naturally leads to a second major theme: variation. Differences in political context are one type of variation; as discussed earlier, political differences in countries' approach to ICTs play a critical role in the articles by Goldstein and Greitens, both because the policies themselves vary and because they create patterns of variation in ICT-derived data. The contributions included here, however, also deal more broadly with patterns of variation within ICT data, and raise the question of how best to use those variations to gain analytical leverage on important questions in political science. Much of the recent public attention to technology's importance in politics has been driven by the supposition that technology has a uniform and prodemocratic impact. Actually, however, it is characterizing and explaining *variations* in the role of technology in political life that will sustain a robust research agenda. We hope, with this symposium, to start a substantive discussion about how to think about the origins of variation in ICT data, its significance for research design, and implications for researching a range of political behavior and outcomes.

A third theme that recurred across contributors was the challenge of validating Internet or ICT data and comparing it to offline behavior. This issue is the central focus of Mellon's piece on how to draw accurate inferences from search data in countries other than the United States. Granka also deals with validation questions in testing the power of search data as compared to more conventional models that predict American election results. Two additional articles, by Koehler-Derrick and Goldstein, focus on using online data to explore offline behavior, but do so in situations where the degree of similarity between the two is more difficult to assess (either because Internet penetration levels are more limited or because official data that could be used as a baseline comparison is unreliable or nonexistent). Validation questions require careful research design, and often some initial testing, to determine what the online data actually reflect. As the contributions here show, the results of those tests can be surprising, and researchers should be careful about the degree to which they assume online data reflect offline behavior. As

https://doi.org/10.1017/S1049096513000255 Published online by Cambridge University Press

the field develops, validation looks likely to remain a significant methodological challenge.

The final theme raised by the symposium contributions was that of scope conditions. We mean that in two senses. First, the articles motivate us to consider the scope conditions of technology as an independent variable. What kind of outcomes do information and communication technologies influence, and what outcomes do they have little effect on? Are there scope conditions that affect which outcomes technology influences, and in what direction?

Second, though, the articles also challenge us to analyze the scope conditions and limitations of ICT-generated data relative to other types of data that could be employed. What is the utility of Internet-derived data in different contexts, and what are its limitations? As political scientists design future research projects, how does Internet- or ICT-derived data compare to, or complement, data obtained through more traditional methods? Several of the contributions here, including those by Granka and Mellon, suggest that search data may have a predictive function, but under conditions that are, as yet, still being explored and which will only be clarified by combining search data with other, more traditional sources. The debate over the potential predictive capacity of ICT-derived data is one example of how the impact and use of ICT data will have to be disaggregated and tested carefully in future work.

CONCLUSION

The data and methodologies derived from the growth of information and communication technologies are both exciting and poorly understood. With this symposium, we seek to create a focal point for conversation among scholars who are doing some of the most innovative work using ICT data. We aim to illustrate the variation in the contexts they draw from, the data they use, and the analytical approaches they apply, and by doing so, to explore not only the common challenges of this emerging body of work, but its potential to make contributions across the discipline of political science. ■

REFERENCES

Bamman, David, Brendan O'Connor, and Noah A. Smith. 2012. "Censorship and Deletion Practices in Chinese Social Media." *FirstMonday* 17 (3). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3943/3169.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489 (7415): 295–98. http://www.nature.com/nature/journal/v489/n7415/pdf/nature11421.pdf.

Carneiro, Herman Anthony, and Eleftherios Mylonakis. 2009. "Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks." *Clinical Infectious Diseases* 49 (10): 1557–64. http://www.ncbi.nlm.nih.gov/pubmed/19845471.

Choi, Hyunyoung, and Hal Varian. 2009. "Predicting the Present with Google Trends." http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/googleblogs/pdfs/google_predicting_the_present.pdf.

Christakis, N., and J. Fowler. 2009. "Social Network Visualizations in Epidemiology." *Norwegian Journal of Epidemiology* 19 (1): 5–16.

Finn, Holly. 2012. "New Gumshoes Go Deep with Data." *The Wall Street Journal*, October 20. http://online.wsj.com/article/SB10000872396390444354004578060654273228618.html.

Ginsberg, Jeremy, Matthew Mohebbi, Rajan Patel, Lynnette Brammer, Mark Smolinski, and Larry Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457: 1012–14.

Jensen, Robert. 2011. "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector." *Quarterly Journal of Economics* 122 (3): 879–924.

King, Gary, Jennifer Pan, and Molly Roberts. 2012. "How Censorship in China Allows Government Criticism but Silences Collective Expression." Paper prepared for the Annual Meeting of the American Political Science Association. New Orleans, August.

Mostak, Todd. 2012. "Social Media as Passive Polling: Using Twitter and Online Forums to Map Islamism in Egypt." MA Thesis, Center for Middle Eastern Studies.

Nadav, Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. "Social fMRI, investigating and shaping social mechanism in the real world." *Pervasive and Mobile Computing* 7: 643–659.

Nielsen, Rich. 2012. "Jihadi Radicalization of Muslim Clerics." Working paper. August 31. Online at http://people.fas.harvard.edu/~rnielsen/jihad.pdf.

Shapiro, Jacob N., and Nils B. Weidmann. 2012. "Is the Phone Mightier Than the Sword? Cell Phones and Insurgent Violence in Iraq." Working paper, September 3, 2012. Online at http://www.princeton.edu/~jns/papers/SW_2011_Cell_Phones_Insurgency_06SEP12.pdf.

SYMPOSIUM AUTHORS

Joshua Goldstein *is a second-year doctoral candidate in the science, technology, and environmental policy (STEP) track at Princeton University. He is interested in using the tools of urban economics, computer science, and psychology to study the link between transparency and governance in developing countries. With a regional focus on sub-Saharan Africa, he has worked extensively on technology policy and open development projects with the World Bank Group, the United States government, Google Inc., and UNICEF. He can be reached at goldstein.joshua@gmail.com.*

Laura Granka *works at Google developing research strategy, conducting research, and managing the team of UX researchers working in the social space. Laura completed her PhD at Stanford University, where she studied selective attention and consumption of public affairs content. She can be reached at laura.granka@gmail.com.*

Sheena Chestnut Greitens *is a doctoral candidate in the department of government at Harvard University, where her work focuses on security studies and authoritarian government in Asia. Beginning in July 2013, she will be an Academy Scholar at the Harvard Academy for International and Area Studies and an assistant professor of political science at the University of Missouri. Greitens can be reached at chestnut@fas.harvard.edu.*

Navid Hassanpour *is a doctoral candidate in the political science department at Yale University, affiliated with the program on order, conflict, and violence (OCV), and a Fellow of the Information Society Project (ISP) at Yale Law School. He can be reached at hassanpour.navid@gmail.com.*

Gabriel Koehler-Derrick *is an associate at the Combating Terrorism Center and an instructor in the department of social sciences at the United States Military Academy at West Point. His research is largely focused on the comparative politics of the Middle East. He can be reached at gabriel.koehler-derrick@usma.edu.*

Jonathan Mellon *is a doctoral candidate in sociology at Nuffield College, University of Oxford. His research focuses on political sociology and measurement in social science, with current research including survey translation, Internet data, and the detection of election fraud. He can be reached at jonathan.mellon@nuffield.ox.ac.uk.*

https://doi.org/10.1017/S1049096513000255 Published online by Cambridge University Press