OPINION

Big data has big potential for applications to climate change adaptation

James D. Ford[a,1,2,3], Simon E. Tilleard[a], Lea Berrang-Ford[a,1,2], Malcolm Araos[a,1], Robbert Biesbroek[b,1,2], Alexandra C. Lesnikowski[a,1,2], Graham K. MacDonald[a], Angel Hsu[c,2], Chen Chen[d,2], and Livia Bizikova[e,2]

The capacity to collect and analyze massive amounts of data is transforming research in the natural and social sciences (1). And yet, the climate change adaptation community has largely overlooked these developments. Here, we examine how "big data" can inform adaptation research and decision-making and outline what's needed from the adaptation community to maximize this opportunity. We contend that careful application of big data could revolutionize our understanding of how to manage the risks of climate change.

Big Data and Climate Change

Although a consistent definition is lacking, there is agreement that big data are characterized by the increasing volume, variety, and velocity of data streams (2, 3). Climate science has long used large datasets to understand the functioning of the climate system, but the field has been slow to use the passively generated information from digital devices and services that characterize big data (4). In the human dimensions community, big data have rarely been used, aside from studies measuring public opinion on climate change based on social media posts. Despite this neglect, vast amounts of geocoded data on human–environment interactions relevant to adaptation already exist digitally and are being added daily. Such data are not limited to the developed world, with widespread global cell phone coverage and social media use for example (5). Potential adaptation applications of these data are numerous.

Vulnerability Assessment

Vulnerability research provides essential information for adaptation decision-making by identifying and characterizing who and what are vulnerable to climate change, to what risks, why, and over what timescales. The vulnerability field has expanded considerably over the last decade, although many gaps in understanding vulnerability processes and drivers remain—particularly concerning the real-time nature of human–environment interactions in a changing climate, absence of longitudinal data, and climate information gaps (6, 7). Big data analytics can help to fill these gaps and have the advantage of being able to use automatically collected real-time data, an important consideration in data-poor environments, especially given the time-consuming nature of survey and interview-based methodologies. Anonymized records of cell-phone use (known as call detail records), for example, allow the large-scale tracing of people's movements, and can be used to examine the dynamics of human mobility in a changing climate (8): Who migrates, to where, and for how long before, during, and after a climate disaster? What are the short- and long-term features of migration behavior?

Anonymized records of cell-phone use could in principle enable the large-scale tracing of people's movements in the wake of a climate change-related disaster. Image courtesy of Shutterstock/Athi Aachawaradt.

[a]Department of Geography, McGill University, Montreal, QC, Canada H3A 2K6; [b]Department of Social Sciences, Public Administration and Policy Group, Wageningen University, 6708 PB Wagnenigen, The Netherlands; [c]Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT 06520; [d]Notre Dame Global Adaptation Index, University of Notre Dame, Notre Dame, IN 46556; and [e]International Institute for Sustainable Development, Ottawa, QC, Canada K1P 5Z9

Author contributions: J.D.F., S.E.T., L.B.-F., M.A., R.B., A.C.L., G.K.M., A.H., C.C., and L.B. wrote the paper.

The authors declare no conflict of interest.

Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and have not been endorsed by the National Academy of Sciences.

[1]J.D.F., L.B.-F., M.A., R.B., and A.C.L. are members of the Tracking Adaptation to Climate Change Collaboration (TRAC3).

[2]J.D.F., L.B.-F., R.B., A.C.L., A.H., C.C., and L.B. are members of the Adaptation Tracking Collaborative.

[3]To whom correspondence should be addressed. Email: james.ford@mcgill.ca.

Big data can also be used to create georeferenced datasets on factors affecting vulnerability–such as population, habitation characteristics, economic status, location of sensitive infrastructure, trends for environmental conditions–which are often lacking, outdated, patchy, or unreliable, especially in low- and middle-income nations (6, 9). Call detail records and geospatial big data, such as high-resolution remote-sensing imagery, allow for dynamic population mapping that can be used to assess risk to natural disasters, characterize habitation patterns of high-risk regions, and track trends over time. Natural language processing techniques can be used to mine and analyze large volumes of text on how climate change discourse is evolving on social media to catalog "bottom-up" perspectives on various components of vulnerability: What are the most frequently talked about climatic risks? Are certain locations consistently identified as being vulnerable in online discussion?

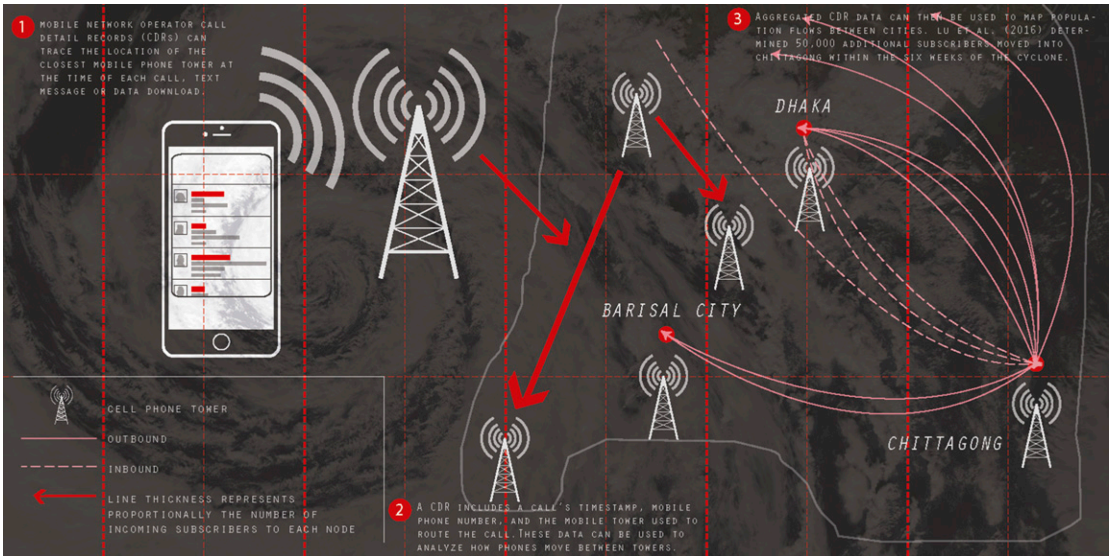
Early Warning

Surveillance and the provision of early warnings is an important component of enhancing the capacity to respond to climate change. Many big data applications have been pioneered for use in early detection, whereby passively collected data from the use of digital services have been variously used: from detecting influenza epidemics based on flu-related queries coming into search engines, to the use of Twitter posts to identify areas affected by earthquakes. Passively collected digital data have the potential to enhance the monitoring of climate-related threats and vulnerabilities, and can provide real-time awareness and feedback to decision makers and emergency services. Hazard warning systems, for example, could incorporate social media data to trigger emergency response measures (e.g., heat or flood alert systems); personal devices equipped with sensors could allow the monitoring of human movement before, during, and after a hazard event to aid with disaster response; tweets can be geotagged so disaster management services can map impacted areas in real-time to target efforts, and the Internet can be scraped for recently uploaded photos of affected areas (10, 11). Search queries could be analyzed to monitor health-seeking behavior to detect outbreaks of climate-related diseases, and changes in the magnitude and frequency of climatic risks could be detected through time-series analysis of multiscale data, with the potential to detect "leading indicators" of abrupt, nonlinear change (12).

Monitoring and Evaluation

Monitoring and evaluating adaptation is methodologically complex, a task made more difficult because of limited data on adaptation actions and outcomes. The result is studies that primarily focus on the process through which interventions are developed and implemented (13). Big data can complement such work by providing metrics on how adaptations actually affect perceptions and behavior.

Cell phone or social media data could be used to examine how an adaptation program designed to provide storm warnings actually affects people's movement before, during, and after a storm event, and monitor if and how this changes over time. Researchers in the aftermath of Hurricane Sandy found the use of Twitter data to be more effective in predicting the location and severity of storm damage than models developed by the US government's Federal Emergency Management Agency (11). Bottom-up perspectives on adaptation development, implementation, and effectiveness can be documented from social media through sentiment analysis of large volumes of posts or by examining "Likes" or "Dislikes" of posts. Building on the success of citizen science initiatives, such as the Massachusetts Institute of Technology's Climate CoLab (a crowdsourcing platform where people work with experts and each other to create, analyze, and select detailed proposals for what to do about climate change, climatecolab.org/) or E-bird (crowdsourcing bird



Lu et al. (8) used data from 6 million mobile phone users to track mobility patterns in Bangladesh following the May 2013 Cyclone Mahasen disaster. Image courtesy of Shreya Shah (Yale University, New Haven, CT).

abundance and distribution, ebird.org/content/ebird), crowdsourcing approaches can be expanded to gather and track information on adaptation policy on the ground.

Efforts to track adaptation across nations or regions have been challenged by an absence of comprehensive, reliable, and up-to-date datasets. Big data can bolster the ability for monitoring environmental change and assessing risk at regional and global scales, with important adaptation applications for climate-vulnerable sectors, such as agriculture. Review of existing research suggests that farmers in some regions are adapting to climate change, for example, by adjusting crop-sowing and cultivation dates to account for changing growing-season conditions (14). Efforts to fuse different satellite remote-sensing products with crowdsourced data have helped to improve the accuracy of global cropland maps (15), and such techniques could be extended to improve assessment of cropping-adaptation activities that are important to farmer livelihoods and food security in low-income and subsistence settings. Predictive analytics could further help farmers to plan and cope with increasing climate variability with location-specific information.

Other opportunities to obtain data for monitoring and evaluation include scraping websites for information on adaptation activities, mining social media to identify and track examples of adaptation, and using financial transaction data of institutions to document adaptation-related financial flows (16). For vulnerable locations, web-scraping methods could be used to retrieve photos posted on social media to see if interventions have prevented or moderated climate-induced processes. The recent launch of Pokemon Go reflects the emergence of technologies that will bring a wide range of people and places into the visual, virtual, and big data domain, with implications for the "gamification" of big data and adaptation of research opportunities.

Combining "Big" and "Small" Data

The big data potential for adaptation research is sizeable, but caution is also warranted. An overreliance on correlations in big data, at the expense of actual causation, can lead to spurious insights (17, 18). Huge datasets provide a high degree of statistical power, but meaningful signals can be masked or lost as researchers simplify and reduce datasets to make them more manageable (19, 20). In the adaptation examples above, for instance, simply tracking social media data density of tweets or social shares in an area pre-/postdisaster doesn't reveal motivations or drivers of migration behavior, and only captures the opinions of those actively engaged with social media. Similarly, analysis of cell-phone data alone holds limited explanatory power on why people move the way they do during extreme events, doing little to advance our understanding of potential future mobility patterns.

The value of big data are thus contingent on our ability to analyze and interpret large datasets in ways that are meaningful and rigorous, and most informative when combined with "small data" approaches. There are two dominant approaches to combining big and small data. A "data-driven" approach uses big datasets to explore—often heuristically, through inductive discovery and investigation, and frequently without a priori hypotheses—emergent patterns and trends in data sources, informing the development of small data studies to investigate observed patterns. A "theory-driven" approach begins with adaptation theory and case studies to identify hypotheses that can be tested with big datasets, seeking to identify generalizable insights and empirically test prevailing theories.

In the adaptation field, many view data-mining approaches and attempts to glean generalizable insights as contemptuous; some argue that these approaches compromise contextual complexity and are ill-suited to adaptation policy questions (16). The adaptation community instead often prefers small-scale studies, which are largely composed of academics and practitioners from a qualitative background. Hence, the top priority is developing in-depth, context-specific insights on adaptation (16, 21, 22).

But it would be big mistake to dismiss big data approaches. For example, big data analyses can identify trends and certain locations that could be targeted for the design of qualitative ground-based surveys or interviews to understand adaptation choices and responses (e.g., rebuild or relocate). Pairing big data insights with in-depth qualitative research can reveal deeper understanding of climate vulnerability, better inform early warning systems, and underpin more rigorous monitoring and evaluation, ultimately leading to more robust adaptation responses. Triangulation of methods and insights is a very powerful tool, yet adaptation research is currently lopsided and dominated by small data.

Big data are no panacea, but if carefully used, they provide an enormous and untapped opportunity to diversify our understanding of adaptation and inform decision-making.

Data access remains a concern. Many datasets are privately held, and individuals are increasingly watchful about how their data are used for commercial and security reasons. Ethical guidelines will be needed to inform how big data are used in adaptation work. Meanwhile, the rise of open-source data, programming languages (e.g., R and Python), and analytic platforms (cloud and cluster computing) have made analysis of big data increasingly accessible and interactive. Although ethical and proprietary concerns will hinder the use of privately held data sources, adaptation researchers should look to open-source platforms and communities for near-term opportunities to harness the potential of big data (e.g., crowdsourcing, coding literacy for adaptation).

Next Steps

Big data are no panacea, but if carefully used, they provide an enormous and untapped opportunity to diversify our understanding of adaptation and inform decision-making. Embracing this opportunity requires a willingness to work with vast amounts of data alongside traditional smaller sampled datasets, appreciation of big data's real world messiness, and the utilization of techniques not typically associated with adaptation work, including data mining, data-driven inquiry, machine learning, and natural language processing. Right now, the adaptation community is poorly equipped to take on this challenge. Scientists investigating adaptation must reach out to data scientists who have the skills to clean, organize, link, manage, and analyze massive datasets; adaptation specialists must work with data engineers to design information systems to scrape, collect, and sort data. In the longer term, more diversified training of adaptation scholars and practitioners to be literate in big data analytics is required.

Funding bodies have a central role in promoting the use of big data for adaptation. Already, some have sought to catalyze the interest of the climate change field, including the United Nation's Global Pulse Big Data Climate Challenge and the White House Climate Data Initiative. These are important steps, but efforts have focused mostly on climate mitigation or making datasets available. In an adaptation context, a starting point for funders could be to invest in demonstration projects that bring together data scientists, adaptation researchers, and decision makers to examine the feasibility of big data approaches and illustrate what's possible. The examples presented herein provide a guide for the kinds of adaptation questions that could be explored through such demonstration projects. Big data also present an opportunity for the private sector because of an increasing demand for information on adaptation needs, options, and services. The challenges of using big data for adaptation applications are significant, but so is the potential. Engaging in big data approaches for adaptation is an opportunity gap waiting to be filled.

Acknowledgments

We acknowledge funding support from the Social Sciences and Humanities Research Council of Canada and the Canadian Institutes of Health Research.

1 Lazer D, et al. (2009) Social science. Computational social science. *Science* 323(5915):721–723.
2 Jagadish HV (2015) Big data and science: Myths and reality. *Big Data Research* 2(2):49–52.
3 Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. Available at https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Accessed September 1, 2016.
4 Faghmous JHK, Kumar V (2014) A big data guide to understanding climate change: The case for theory-guided data science. *Big Data* 2(3):155–163.
5 James J (2016) *Data never sleeps 3.0.* Available at https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/. Accessed September 1, 2016.
6 Moss RH, et al. (2013) Climate change. Hell and high water: Practice-relevant adaptation science. *Science* 342(6159):696–698.
7 McDowell G, Ford J, Jones J (2016) Community-level climate change vulnerability research: Trends, progress, and future directions. *Environ Res Lett* 11(3):033001.
8 Lu X, et al. (2016) Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Glob Environ Change* 38:1–7.
9 Hsu A, Malik O, Johnson L, Esty DC (2014) Development: Mobilize citizens to track sustainability. *Nature* 508(7494):33–35.
10 Keeler BL, et al. (2015) Recreational demand for clean water: Evidence from geotagged photographs by visitors to lakes. *Front Ecol Environ* 13(2):76–81.
11 Kryvasheyeu Y, et al. (2016) Rapid assessment of disaster damage using social media activity. *Sci Adv* 2(3):e1500779.
12 Biggs R, Carpenter SR, Brock WA (2009) Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proc Natl Acad Sci USA* 106(3):826–831.
13 Ford JD, et al. (2015) Adaptation tracking for a post-2015 climate agreement. *Nat Clim Chang* 5(11):967–969.
14 IPCC (2014) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds Field CB, et al. (Cambridge Univ Press, Cambridge, UK), pp 1132.
15 Fritz S, et al. (2015) Mapping global cropland and field size. *Glob Change Biol* 21(5):1980–1992.
16 Berrang-Ford L, Pearce T, Ford JD (2015) Systematic review approaches for climate change adaptation research. *Reg Environ Change* 15(5):755–769.
17 Lazer D, Kennedy R, King G, Vespignani A (2014) Big data. The parable of Google Flu: Traps in big data analysis. *Science* 343(6176):1203–1205.
18 Palen L, Anderson KM (2016) Social media research. Crisis informatics—New data for extraordinary times. *Science* 353(6296):224–225.
19 Gandomi A, Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manage* 35(2):137–144.
20 Giardullo P (2016) Does 'bigger' mean 'better'? Pitfalls and shortcuts associated with big data for social research. *Qual Quant* 50(2):529–547.
21 Swart R, Biesbroek R, Lourenco TC (2014) Science of adaptation to climate change and science for adaptation. *Front Environ Sci* 2:29.
22 Javeline D (2014) The most important topic political scientists are not studying: Adapting to climate change. *Perspect Polit* 12(2):420–434.