Online Annotations for

Can We Do Better? Replication and Online Appendices in Political Science

by REDACTED

#1

As Arthur Lupia and Colin Elman (2014, 22) emphasize, openness is essential for an effective dialogue between members of different sub-disciplinary and methodological communities within political science.

Arthur Lupia and Colin Elman (2014, 22) emphasize that openness is essential for an effective dialogue between members of different sub-disciplinary and methodological communities within political science. They explain:

Second, openness is beneficial for scholars outside the immediate community in which the research is located. Political science is a methodologically diverse

discipline, and we are sometimes unable to appreciate how other social scientists generate their conclusions. Mathematical modelers, for example, often know very little about how cases are selected in participant observation studies--and many people who seek meaning in texts have a limited understanding of how other social scientists try to seek meaning from surveys or computer simulations of war. Higher standards of data access and research transparency will make cross-border understanding more attainable.

Other audiences are not focally involved in research. Instead, they want to use research claims as the basis of action. Teachers, for example, want to use the claims for pedagogical purposes. Whether demonstrating substantive arguments about aspects of the social world, or training students to use research techniques, teaching is substantially improved by the availability of exemplary scholarship, with its data and reasoning on display.

Public and private sector decision makers comprise another audience. Their main interest is in using knowledge claims to improve the effectiveness and efficiency of valuable endeavors. Greater openness gives such audiences increased opportunities to understand how the claims relate to their aspirations. As Lupia (2014) notes, many decision makers value information whose veracity they can readily defend in politicized contexts. These decision makers find claims whose origins are available and accessible more valuable informational currency than claims whose foundations are hidden.

#2

Accessibility and user experience design, in turn, are prerequisites for real openness. They can make the difference between primary replication, in which scholars are able to fully replicate the entire course of a study, and secondary replication, which only entails replicating the statistical analysis based on data and code shared by authors.

When discussing solutions to the replication challenge, delineating the boundaries of the concept can be helpful. A narrow definition suggests that replication is the ability to reproduce the analysis that led the authors to their conclusions. Indeed, it is common to find appendices of quantitative studies that include a data file, a brief explanation of the data and variables, output of statistical operations, and preliminary analytical steps that do not appear in the article itself. A broader definition of replication encompasses a detailed account of every step that the researchers have taken after formulating the hypotheses. This is the only way to replicate not only the analysis of the data but also the preceding steps. While this approach offers full transparency, it sets very high standards of documentation. Adopting it is likely to make researchers substitute curiosity, creativity, and exploration with a rigid, risk averting technical adherence to the

process. While we acknowledge this tradeoff, addressing it would require a separate article and is therefore beyond the scope of this one.

#3

Throughout this decade, a series of formal and informal discussions and deliberations concerning transparency took place in political science conferences, committees, journal symposia, and online forums, resulting in various reports and recommendations.

For a concise yet illuminating history of the DA-RT movement, see Golder and Golder 2016, especially the section entitled "A Brief History of DA-RT" (2–5). Additional online resources can be found in the DA-RT section of APSA's *Political Science Now* website.

#4

However, recent technological developments have dismantled these obstacles. Thanks to the steep drop in the cost of digital storage and the emergence of numerous academic and non-academic data archives, researchers and journals can now host and share raw and structured replication data freely or at an affordable cost.

Logistical difficulties, such as the cost of digital stage, have more recently dominated discussions about primary replicability. While the shift to replicability has led to fewer critiques of supplementary information as cumbersome and awkward, the concern still remains that lengthy methodological explanations could disrupt the flow of an article's narrative, destroy the elegance of a text, and occupy too much valuable space. For reference, see Ian Lustick's influential article from 1996, in which he writes (616):

[...]scholars might choose to share the qualitative judgments that led to choices of particular sources for constructing different parts of the background narrative. […] This technique would be supplemented by discursive footnotes that alert the reader to alternative versions and briefly explain the reasoning that led to the rejection of these accounts. [...] One substantial drawback to this strategy, especially in light of constraints imposed by publishing houses, is that both the level of detail and the length of manuscripts could be significantly increased [...]

Similarly, in the conclusion of a more recent piece in *Security Studies*, Nina Tannenwald (2015, 229) discusses a possible drawback of explicit process tracing explanations in security studies:

Security studies is not simply method; it is also good storytelling. In the welcome move to greater self-consciousness and transparency in carrying out process tracing, my hope is that it continues to complement or even enhance, but does not come at the expense of, good narrative.

In the past, such explicit methodological explanations had to be included in the main text, or in a printed appendix that followed it. Today, however, many access academic literature online and often read it from a screen rather than paper. For example, see this 2015 study by Carol Tenopir, Donald W. King, Lisa Christian, and Rachel Volentine, whose abstract reads:

Electronic journals are now the norm for accessing and reading scholarly articles. This article examines scholarly article reading patterns by faculty in five US universities in 2012. Selected findings are also compared to some general trends from studies conducted periodically since 1977. In the 2012 survey, over three quarters (76%) of the scholarly readings were obtained through electronic means and just over half (51%) of readings were read on a screen rather than from a print source or being printed out. Readings from library sources are overwhelmingly from e-sources. The average number of articles read per month was 20.66, with

most articles read by the medical and other sciences, and on average each article was read for 32 minutes.

Consequently, the need to include in an article's main text all the underlying data and information has become obsolete. Instead, such elements can be posted separately and digitally, without narrowing down their scope. In other words, the main article's word limit should not influence the length of the appendix. Whereas the former offers a concise and polished account of the literature reviewed, theories and methodologies used, data analyzed, and conclusions drawn, the supplementary materials allow readers to explore the article's underlying "plumbing system" and follow in the investigator's footsteps throughout the entire analytical process – as numerous, tedious, and complicated as these steps may be. Thus, in some journals, two-page articles can be accompanied by appendices that are dozens of pages long. As a case in point, consider this excellent article on event data by Wei Wang, Ryan Kennedy, David Lazer, and Naren Ramakrishnan (Wang et al. 2016), as well as its appendix.

#5

Out of 79 replication datasets, only 47 include a README file listing the contents of the data repository and/or a codebook explaining how the variables were coded. Many of these documents are concise, laden with technical jargon.

One of the challenges for appendix readability is inaccessible technical jargon. While authors strive for clear language in articles, appendices must also be well-written, free of jargon and obscure language, to be accessible. See Michael Billig's (2013) book *Learn to Write Badly: How to Succeed in the Social Sciences*.

#6

Making our scholarship clearer does not mean that we need to "dumb down" our argument and methodological explanations or abandon our rigor; scholarly quality and accessibility need not be mutually exclusive.

Accessibility can reinforce standards of academic rigor, as Matthew Flinders (2015, 75) asserts:

Critics of my position will undoubtedly argue that political science, as a professional discipline, will inevitably require the use of certain technical terms or phrases that are understandably not within the mainstream public vocabulary. This, again, is rarely more than a smokescreen. Technical terms will, of course, have to be used from time to time, but 'technical' does not necessarily mean difficult, and certainly does not mean jargon. Political science has become a discipline built on jargon; and if technical terms are really necessary and also clear and precise, it is not difficult to use them in the context of plain English and thus introduce them meaningfully to the reader. Critics may at this point engage in a far more sinister and hurtful form of criticism and accuse me of advocating the demotion of academic scholarship into little more than pseudojournalism. This sideswipe will be couched upon the implicit suggestion that I am obviously unable to grasp the intellectual magnitude of their work and am therefore trying to lower the standard of political science towards my own inferior level. The curse of political science, a curse that both Mills and Crick endured, is to become identified as a 'mere literary man' or, worse still, to have their work defined as 'mere journalism'. Any academic who dares to write in a widely intelligible way, let alone engages with television or radio, is liable to be condemned in this manner. This reflects a rather superficial logic. Accessibility and scholarly quality do not exist in a zero-sum relationship whereby an increase in one inevitably leads to a reduction in the other.

Of course, we do not expect that scholars write their explanations to allow readers without the necessary methodological training to carry out the analysis based on the appendix. What we do expect, however, is that scholars in the field will be able to understand what type of information and files can be found in the appendix, how to access these materials, and what programs and methodological skills are necessary to open them and perform the analysis.

#7

As Paul Musgrave and Sebastian Karcher state with respect to online annotations, the unique features of appendices can help us to "showcase the depth" of our research and "make it easier for readers to go beyond the article."

Paul Musgrave and Sebastian Karcher (2018) highlight the additional depth that an appendix can provide:

ATI [Annotations for Transparent Inquiry] annotations allow both author and reader to go deeper in the research process. Is there a complex backstory to obtaining a piece of evidence? Include it in an annotation. Would a longer excerpt

from an interview or document than you have room for in your article bolster your claims? ATI annotations offer unlimited room for such excerpts. Is there a document that would be hard for readers to obtain but provides valuable context? You can include it as a data source.

At the core of ATI is the idea to showcase the depth of qualitative research. Research goes far beyond the 10,000 words of your article, and ATI gives authors a way to showcase this depth—not only to make the work more convincing but also to make it easier for readers to go beyond the article.

In a similar vein, Gary King, Robert Keohane, and Sidney Verba explained in their iconic *Designing Social Inquiry: Scientific Inference in Qualitative Research* (1994, 26) that:

Replicability is important even if no one actually replicates our study. Only by reporting the study in sufficient detail so that it can be replicated is it possible to evaluate the procedures followed and methods used.

#8

Specific words and sentences in the appendix can include hyperlinks to open access resources like Wikipedia entries for key terms, actors, and events.

Hyperlinks in the appendix enable scholars to trace the researcher's process directly to any referenced articles. A good example of such practices is the 2018 freely downloadable, open access book *The UK's Changing Democracy: The 2018 Democratic Audit*, edited by Patrick Dunleavy, Alice Park, and Ros Taylor. While this edited volume does not have an online appendix, it is a digital book meant to be read on a screen. As such, its main text contains direct hyperlinks to (mostly free versions of) studies and evidence rather than traditional references. Such a bibliography is provided at the end of the book and includes the URLs of the cited items. In the preface to the book, the editors explain:

(i) Wherever possible we have sought to link to digitally available sources. Not only this, but when you click on any URL in the text it should also take you to a

completely free and open access source for further information or reading. So far as we possibly can arrange it, whenever you click on our hyperlinks you should never go just to a paywall.

[...]

(ii) Where no freely readable version is feasible for papers or for books we have linked to blogs, press articles or other short pieces where academics outline their work in accessible ways. (Students and academics with university library access can easily move on from there to the full texts.)

[...]

(iii) On factual matters (such as election data) we have linked a great deal to official statistics, and the many invaluable reports and databases from the House of Commons Library and Institute for Government. Our links to Wikipedia also reflect the fact that its coverage is increasingly broad, reliable and up to date, although we have carefully checked the items used here.

#9

The types of data to be included in appendices vary from photos of material objects and ancient manuscripts, to reflexive field notes and interview recordings, to software syntax and output.

To provide all relevant sources in an appendix, researchers can upload independent files to their digital appendices or embed external files in them. The latter option, of course, risks the possibility that, at some point, the owners of these files might remove them, change their location, or prevent access to them. Therefore, it should be generally avoided. However, if we can legally create a copy of these files or parts of them, uploading such copies to our own appendices would be preferable to linking these appendices to another's files elsewhere.

In the case of qualitative research, the files uploaded to appendices may include, among other things, audio and video clips (for example, oral interviews or ethnographic observations); scans of archival documents or survey questionnaires; geospatial data and annotated maps; photographs of field research sites, interviewees, or material sources. Authors of statistical analyses can embed, in their appendices, tables, graphs, and charts along with digital dashboards that allow readers to interact with the underlying data.

#10

To attain these goals, both authors and journals need to take action.

Both authors and journals need to make changes to improve replicability. As John Ishiyama (2014, 78) explains:

On the one hand, the individualistic model holds that the primary responsibility for making data available for replication purposes lies with the individual author. On the other hand, the social or community policy makes the provision of replication data part of the publication process (and required by journals as requirement for publication). In this case the journal, as a representative of the scholarly community, is the responsible to make sure that data for replication purposes is provided to that community.

#11

A political scientist, even without familiarity with the author's methodology or field of expertise, should be able to navigate the appendix, access the files, and understand their nature, purpose, and the type of software needed to open them.

For a good source about user experience design, with many practical guidelines and recommendations that are applicable to online appendices, see this book by Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven M. Jacobs, and Niklas Elmqvist (2018). They advise:

- ***Use unique and descriptive headings*. Use headings that are distinct from one another and conceptually related to the content they describe (page 84).**

- ***Consistency of data display*. During the design process, the terminology, abbreviations, formats, colors, capitalization, and so on should all be standardized and controlled by use of a dictionary of these items (page 85).**

- ***Seek universal usability*. Recognize the needs of diverse users and design for *plasticity*, facilitating transformation of content. Novice to expert differences, age ranges, disabilities, international variations, and technological diversity each enrich the spectrum of requirements that guides design. Adding features for novices, such as explanations, and features for experts, such as shortcuts and faster pacing, enriches the interface design and improves perceived quality (pages 95–96).**

#12

We should attempt to use common file formats that can be opened with freely available software – for example, PDF, TXT, or RTF file extensions for text documents or CSV files for structured data. Files that require specialized software should be uploaded alongside, and not in place of, more common file types.

We model our proposals in the online appendix to this article by sharing data in common file types as well as files requiring specialized software. See our dataset in its original form, in the spreadsheet application *Airtable* used to create it. In addition, we offer the raw data as a downloadable CSV file that can be opened in many types of software.

#13

To ensure integrity, every appendix should have an "anchor" – a web page or text document serving as the roadmap for the entire appendix. Like the table of contents in a text document or the README file in a data repository, this page should utilize clear language and link to every section or file in the appendix.

R. Michael Alvarez, Ellen M. Key, and Lucas Núñez (2018) recommend that README files of replication data include the following elements:

a) A reference to the associated paper or publication.

b) A short description of the files and file types included in the replication package: for example, raw data, processed data, scripts to manage the data, and scripts to produce estimates.

c) An indication of the order in which the scripts are to be run, as well as noting where the different tables and figures found in the associated publication are generated and stored. [...]

d) A list of software and software packages (as well as the dependent packages on which they rely) and the operating system used to produce results in the paper. Technical information on the hardware used to produce the results also is helpful (e.g., the number of cores), especially if computationally demanding techniques were applied. [...]

e) If unusual file extensions are used, authors should clearly indicate—to the extent possible—how to proceed with these files.

Also, see here an interesting discussion of README files in code development, including good advice on how to create effective READMEs.

#14

To implement principles of integrity and intuitiveness in appendices, every journal needs to adopt some standards regarding the way authors create, store, display, and format their appendices.

For political scientists to create better appendices, the practice of writing and developing appendices should be taught. In addition to academic writing courses, higher education institutions should offer courses that instruct graduate students in crafting appendices. Currently, researchers must acquire these skills during their training while interacting with advisors and other experienced scholars. Since methodological appendices for replication are a relatively new requirement for social science researchers, putting such files together is a skill that even veteran researchers may not have. To ensure that all researchers develop these skills, professional associations and journals should offer detailed guidelines and workshops for the preparation of appendices to enable replication.

#15

Authors of qualitative appendices should describe their selection of sources, methodological and analytical choices, and hypotheses tested (including ones that were rejected). They should also aim at exposing readers to the raw data – documents, recorded interviews, and other sources or parts of them – that led to their conclusions, insofar as these materials can be legally and ethically shared.

Ideally, every appendix in the qualitative domain should provide direct access from the article to all documents, recorded interviews, and other sources that led the authors to their conclusions. Additionally, if it does not appear in the article itself, the researcher should offer his or her reasoning for any conclusions and the tools that she or he used to reject alternative interpretations. Of course, such perfect replicability is usually impossible in qualitative or quantitative studies. As even the most enthusiastic adherents of research transparency concede, transparency requirements always "defer to intellectual property, human subject, and logistical constraints" (Moravcsik 2014, 686).

The legal scholar Steven Lubet offers creative advice on how to deal with such limitations in the conclusion of *Interrogating Ethnography: Why Evidence Matters*, his 2018 book on the recent replication crisis in ethnography. Computer-Assisted Qualitative Data Analysis Software

(CAQDAS) such as *NVivo* also produce output files that can be shared in online appendices.

In quantitative research, the main text document of the appendix should serve as a detailed roadmap of the entire research process, from the choice of a research method to the analysis and interpretation of the results. This document would offer a level of detail that is not possible within the corresponding section of the article itself, yet imperative for replication purposes. Among other things, the document should include a discussion explaining the choice of methods, a detailed description of the research population, and preliminary analytical steps that the researchers have conducted but do not merit inclusion in the article itself.

The other essential parts of the appendix of a quantitative article are the database, codebook, and analytical script. The script (or syntax) refers to the code lines that represent each step of data preparation and analysis. Researchers who prefer graphical user interface (GUI) based software, such as the proprietary *SPSS* or the free *PSPP*, should save the automatically generated syntax that the software produces in the background. In command-line interface (CLI) software like *R*, the script file represents an aggregation of all commands entered by the investigator.

#16

While some journals have already developed rather detailed instructions for creating appendices, these standards need to be better enforced and more widely implemented.

Some political science journals have detailed guidelines for appendices. For example, the American Journal of Political Science's guidelines for manuscript preparation regulates the number of pages of an appendix's text document, as well as its formatting. Among other things, the guidelines stress that "All SI [supporting information] files should have a title page and a table of contents identifying the content by page number; this content and the specific page numbers must be referred to at the appropriate place in the main text of the manuscript."

Some journals outside the discipline have even more rigid rules for creating appendices. For example, *Science* not only offers very detailed instructions regarding the "Format and Style of Supplementary Materials," it actually allows authors to download a Microsoft Word template of an appendix. As for the content of the appendix, *Science* clarifies that "Further discussion or development of arguments beyond those in the main text is not permitted in

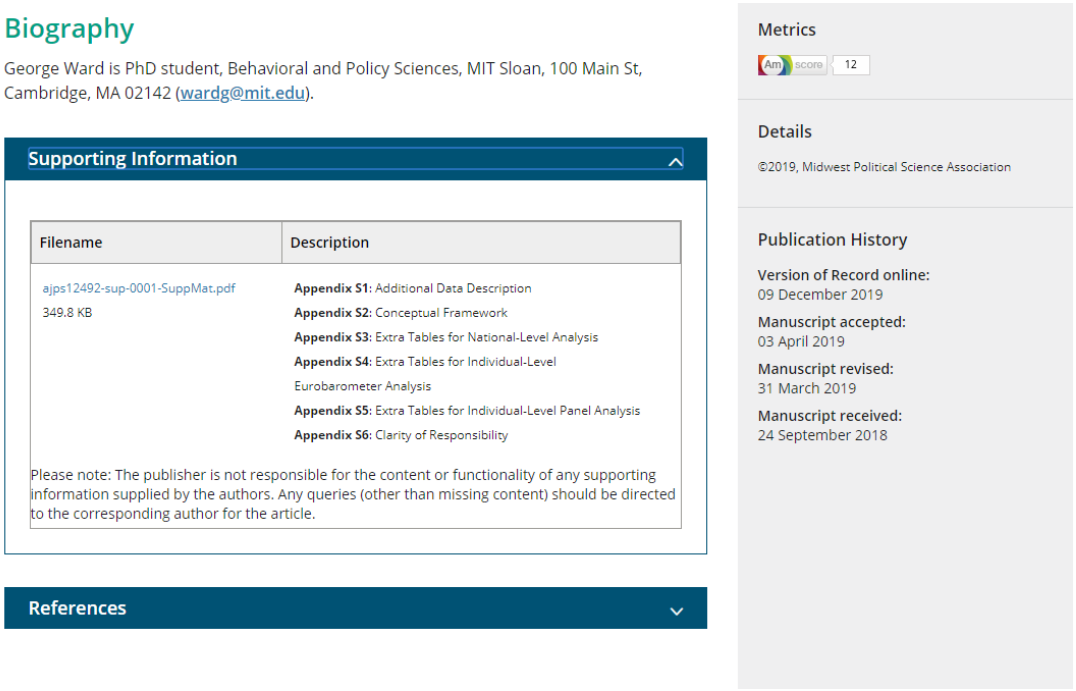
supplementary text." This requirement stands in contrast with the conceptualization of the appendix as a possible means to expand the scope of the article, as argued, for example, by proponents of online annotations. Whether the potential of such stringent measures to enable better replication is worth the risk of stifling creativity and originality is a question that, at some point, the discipline will have to address.

#17

Given the high costs of such an endeavor, journals may be unable to take on the maintenance of appendices themselves.

As a case in point, consider the American Journal of Political Science (AJPS), which makes the following disclaimer with regard to online appendices:

"Please note: The publisher is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing content) should be directed to the corresponding author for the article."



Biography

George Ward is PhD student, Behavioral and Policy Sciences, MIT Sloan, 100 Main St, Cambridge, MA 02142 (wardg@mit.edu).

Supporting Information

Filename	Description
ajps12492-sup-0001-SuppMat.pdf 349.8 KB	**Appendix S1:** Additional Data Description **Appendix S2:** Conceptual Framework **Appendix S3:** Extra Tables for National-Level Analysis **Appendix S4:** Extra Tables for Individual-Level Eurobarometer Analysis **Appendix S5:** Extra Tables for Individual-Level Panel Analysis **Appendix S6:** Clarity of Responsibility

Please note: The publisher is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing content) should be directed to the corresponding author for the article.

References

Metrics

Details

Publication History

Version of Record online: 09 December 2019

Manuscript accepted: 03 April 2019

Manuscript revised: 31 March 2019

Manuscript received: 24 September 2018

Screenshot taken 16 December 2019 from the web page of George Ward's article "Happiness and Voting: Evidence from Four Decades of Elections in Europe."

#18

The editors can verify the existence of an appendix as well as its adherence to the journal's requirements only prior to publication. Once an article is published, it would be almost impossible to enforce authors to keep appendices accessible and updated.

One possible solution for the challenges to updating appendices would be that journals cooperate with CMS providers, so that once an appendix is

"sealed," the journal can transfer it to its servers, host it there, and thus have control over it. Another possibility is that journals and CMS providers would offer authors the tools to create such appendices that would be hosted on the journal's page.

References for the Annotations

Billig, Michael. 2013. *Learn to Write Badly: How to Succeed in the Social Sciences*. Cambridge ; New York: Cambridge University Press.

Dunleavy, Patrick, Alice Park, and Ros Taylor, eds. 2018. *The UK's Changing Democracy: The 2018 Democratic Audit*. London: LSE Press.

Flinders, Matthew. 2015. "The Rediscovery of the Political Imagination." In *The Relevance of Political Science*, Political analysis, eds. Gerry Stoker, B. Guy Peters, and Jon Pierre. Basingstoke ; New York: Palgrave, 65–83.

Golder, Matt, and Sona N. Golder. 2016. "Symposium: Data Access and Research Transparency (DA-RT)." *Comparative Politics Newsletter* 26(1): 2–24.

Ishiyama, John. 2014. "Replication, Research Transparency, and Journal Publications: Individualism, Community Models, and the Future of Replication Studies." *PS: Political Science & Politics* 47(1): 78–83.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, N.J: Princeton University Press.

Lubet, Steven. 2018. *Interrogating Ethnography: Why Evidence Matters*. New York: Oxford University Press.

Lupia, Arthur. 2014. "What Is the Value of Social Science? Challenges for Researchers and Government Funders." *PS: Political Science & Politics* 47(1): 1–7.

Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency: Introduction." *PS: Political Science & Politics* 47(1): 19–42.

Lustick, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90(3): 605–18.

Moravcsik, Andrew. 2014. "Trust, but Verify: The Transparency Revolution and Qualitative International Relations." *Security Studies* 23(4): 663–88.

Shneiderman, Ben et al. 2018. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Sixth Edition. Boston: Pearson.

Tannenwald, Nina. 2015. "Process Tracing and Security Studies." *Security Studies* 24(2): 219–27.

Tenopir, Carol, Donald W. King, Lisa Christian, and Rachel Volentine. 2015. "Scholarly Article Seeking, Reading, and Use: A Continuing Evolution

from Print to Electronic in the Sciences and Social Sciences." *Learned Publishing* 28(2): 93–105.

Wang, Wei, Ryan Kennedy, David Lazer, and Naren Ramakrishnan. 2016. "Growing Pains for Global Monitoring of Societal Events." *Science* 353(6307): 1502–3.