Review

Bigger Data and Quantitative Methods in the Study of Socio-Environmental Conflicts

Paul Alexander Haslam
School of International Development and Global Studies, University of Ottawa, Ottawa, ON K1N6N5, Canada; phaslam@uottawa.ca

Received: 18 August 2020; Accepted: 15 September 2020; Published: 17 September 2020

Abstract: New data sources that I characterize as “bigger data” can offer insight into the causes and consequences of socio-environmental conflicts, especially in the mining and extractive sectors, improving the accuracy and generalizability of findings. This article considers several contemporary methods for generating, compiling, and structuring data including geographic information system (GIS) data, and protest event analysis (PEA). Methodologies based on the use of bigger data and quantitative methods can complement, challenge, and even substitute for findings from the qualitative literature. A review of the literature shows that a particularly promising approach is to combine multiple sources of data to analyze complex problems. Moreover, such approaches permit the researcher to conduct methodologically rigorous desk-based research that is suited to areas with difficult field conditions or restricted access, and is especially relevant in a pandemic and post-pandemic context in which the ability to conduct field research is constrained.

Keywords: big data; socio-environmental conflicts; mining; geographic information systems (GIS); protest event analysis (PEA); social media; research methods

1. Introduction

The use and analysis of bigger data is the latest game changer in the social sciences but is only beginning to be applied in the social scientific study of socio-environmental conflicts. The vast majority of the literature on social conflicts related to resource mega-projects has been conducted using traditional case study methods and intensive fieldwork. In this regard, there are great opportunities to complement or strengthen these analyses through the use of quantified data derived from a wide range of sources including remote (satellite) sensing, academic databases, NGOs, and users of social media.

The purpose of this article is to survey the uses of bigger data in the literature on socio-environmental conflicts to date, and also to hypothesize about the potential of some relatively unexploited sources of data in this subfield. I begin by defining “bigger data”, especially in comparison to the commonly used term “big data”, and place the debate about its use and utility within the larger context of the methodological advantages and limits of different research techniques. In the second part, I examine the application of bigger data to socio-environmental conflicts focusing on geographic information system (GIS) analysis, and protest event analysis (PEA). I conclude with a brief discussion of the ethical implications of research using bigger data.

Throughout the article, I emphasize the ways that alternative sources of data can open new pathways to understanding the complexity of socio-environmental conflicts and providing rigorous evidentiary support for hypotheses generated by qualitative research. Fundamentally, this article is an argument about the potential of data that is hidden in plain sight to transform the way we analyze socio-environmental conflicts, and is addressed particularly to qualitatively-oriented researchers. However, it is beyond the scope of this paper to address the technical issues of how to conduct a quantitative analysis using bigger data.
2. What Is “Bigger Data” and What Is Its Value?

If qualitative case studies and the intensive engagement of the researcher with human subjects over a relatively long period of time can be characterized as small data, then big data implies a vast amount of data, collected remotely and impersonally. The textbook definition of big data distinguishes it from regular data by virtue of its velocity, volume, and variety, and underlines that its analysis requires advanced quantitative and programming skills [1,2]. For the most part, the raw data that comprise big data are generated by the interaction of individuals with the modern economy, which creates an electronic record of every transaction; although remote-sensing technologies (such as satellites) also create continuous streams of data that are independent of human interactions, and which may also be considered to be “big” [2] p. 4. While big data is often invoked as a commercial opportunity for private companies, perhaps exemplified by Cambridge Analytica’s notorious exploitation of a Facebook application to collect data that it subsequently used for micro-targeted political campaigns, it has also sparked a “big data for development” (BD4D) movement that focuses on the use of data to understand, solve, and predict development problems [2,3].

In this article, I focus on “bigger data” rather than “big data”. By making this semantic elision, I wish to underline the similar opportunities now available to exploit “new” sources of data, but also to recognize important differences, particularly along the dimension of the velocity of (real-time) data, which is less likely to be exploitable by social scientific researchers (although not impossible). Bigger data, therefore, refers to quantified or quantifiable data from varied sources that can be collected in large volumes and which require quantitative analysis. The concept of bigger data allows us to recognize that in the social sciences a lot of useful data still needs to be processed (extracted, quantified, and organized) by researchers, and that it is the combination of this data with other sources, some of it legitimately “big” like remote sensing data, which can provide really useful leverage over problems. Bigger data can be seen as an intermediate type of data, falling between small data (case studies) and big data (see Table 1). In this regard, what I am calling bigger data is similar to traditional approaches to quantification, except that I want to explicitly emphasize the opportunities deriving from relatively new sources of information and new combinations of data that can be used in the study of socio-environmental conflicts.

| Characteristics of data collection | Small Data | Bigger Data | Big Data |
|-----------------------------------|------------|-------------|----------|
| Engagement with research subjects | In situ (fieldwork) | Remotely | Impersonal |
| Velocity of data capture | Slow (based on field observations and interviews) | Medium (labor-intensive processing of data prior to analysis) | Rapid (real-time or close to it) |
| Volume of data captured | Limited units of analysis but highly varied content | Intermediate number of units of analysis; varied content | Unlimited units of analysis but narrowly defined content (based on instrument providing the data) |
| Variety of data captured | Qualitative and labor-intensive | Quantitative and labor-intensive | Real-time quantitative (automated analysis) |
| Analysis | Contextualized interpretation of individual actions and motivations | Novel combinations of data improve validity and generalizability of findings | Real-time description of changes in individual behavior |
| Analytical advantages | Weak validity and generalizability | Time-consuming data processing for analysis; quantitative analysis skills required; conceptual stretching often required | Access to data highly restricted (due to privacy and ownership issues); technically complex to process |
| Analytical disadvantages | | | |

Table 1. Comparison of Small Data, Bigger Data, and Big Data.
Sustainability 2020, 12, 7673 3 of 12

Socio-environmental conflicts involve the interaction of firms, community groups, and various institutions of the state, with a particular focus on livelihood and distributional questions. In this regard, the sources of bigger data principally derive from the interaction of these actors with the public sphere. Firms provide data to stakeholders, shareholders, and regulatory agencies (CSR reports, annual reports, stock market and environmental impact assessment filings); community groups seek to cultivate public support (postings on social media, networking with other NGOs, reporting by traditional media); and states increasingly provide public information as part of transparency initiatives (financial data, ombudsman reports, environmental evaluation, censuses, and sector survey results). Satellite remote sensing data covering land-use patterns are also available through various sources, and increasingly governments are seeking to provide some of the data mentioned above in a geo-referenced format (such as population and poverty data from censuses). Some of this data is available in an aggregated and structured format suitable for quantitative analysis, but much of it has to be painstakingly extracted and coded (converted from descriptive text) by the researcher.

Methods in the Study of Social Conflicts

For the most part, the advantages and limits of using bigger data over case studies are the same as traditional quantitative methods. It is well-known that case studies have certain advantages including unparalleled descriptive richness, exploring complex relationships, conceptual validity (no need to dumb down concepts or use conceptually-stretched proxy variables as is common in quantitative work), deriving new hypotheses, and exploring causal mechanisms [4–6]. On the other hand, single cases in particular risk being purely descriptive, casually indeterminate, and not generalizable, except under certain narrow conditions such as being a representative instance of their class, or when cases can be classified as “most-likely” or “least-likely” and twinned with the most appropriate scientific strategy—respectively, falsification and confirmation. Most importantly for the generation of widely applicable theoretical insights, Lipjhart observes “a single case can constitute neither the basis for a valid generalization nor the ground for disproving an established generalization” [7,8].

A serious methodological problem that affects a lot of case study research into socio-environmental conflicts is the study of emblematic cases, which creates “selection bias”. This constitutes selecting a case for study based on the value of the dependent variable, meaning the outcome of interest is known (like existence of a social conflict). This is a common trap for case-study research where such cases can be exciting and attractive, supported by more secondary literature, media reports, and activist engagement. Such cases can also be “interesting” to colleagues and funders because they are known. However, King, Keohane, and Verba view this as a cardinal sin of methodology, arguing that when “observations are selected on the basis of a particular value of the dependent variable, nothing whatsoever can be learned about the causes of the dependent variable” from that case study [5] (p. 129). Where cases are selected because they are emblematic, they also exhibit extreme values (by definition), which contribute to the overestimation of causal effects [5] (p. 139). Rigorous case selection and the use of comparative between-case methods, such as George’s “structured, focused comparison” [4,9], can help address these problems within the case study approach as demonstrated in the exemplary comparative methodology of Amengual (2018) and Steinberg (2019) [10,11].

Another potential methodological risk in case-study work is subjective (researcher) bias. Subjective bias should be distinguished from case-selection bias, although the two are related. Most researchers working on socio-environmental conflicts see themselves as engaged scholars who are understandably more attuned to activist than company perspectives, and such natural affinities are frequently compounded by the reality that access to activists is often conditional on support for their activities. As a result, activist perspectives may be uncritically presented in research findings. Objective (quantified) data on company and community characteristics, protest events, land-use patterns, and institutional actions can offer a corrective to researcher bias.

Of course, quantitative studies can also suffer from selection and subjective bias, and such an approach is typically found in descriptive statistics that count-up the number of conflicts or allegations.
of human rights abuses, without consideration of no-conflict cases [12,13]. If the sample is constituted by conflict cases, the researcher cannot determine the causes of conflict statistically, and must instead focus on some variation within the sample such as activist strategies or outcomes of activism [14]. For the most part, the advantages and limits of using bigger data over case studies are the same as traditional quantitative methods. If a sample of cases is drawn randomly (or universally in the case of user-generated big data), then the use of statistical methods and bigger data has the potential to address the most important weaknesses of case study research: selection bias effects and the lack of generalizability. In the case of genuine big data sources, high velocity (close to real-time) data, including remote sensing and satellite imagery, call-log and locational data from mobile phones, or social media feeds, can allow researchers to understand problems rapidly, short-cutting the months and years required to generate qualitative data and analysis [2] (p. 6).

The issues of researcher bias, selection bias, and generalizability are fundamental to our ability to know and say something about the social world with a significant degree of confidence. In this regard, the promise of bigger data to improve the quality findings in the study of social conflicts is very significant. At a minimum, the use of alternative quantified sources of data can contribute to the triangulation of findings, allowing rigorous testing of hypotheses generated qualitatively.

3. Approaches to the Inclusion of Bigger Data in the Study of Socio-Environmental Conflicts

The analysis that opened new vistas on the use of bigger data to understand socio-environmental conflicts was Javier Arellano-Yanguas’ 2011 article on the local resource curse [15]. His analysis used “new” sources of publicly available data; applied a statistical analysis for the first time in a widely cited analysis; and pushed the quantitative analysis to the subnational level (Peruvian departments equivalent to states or provinces in federal countries). In this case, Arellano-Yanguas was able to leverage the newfound transparency of the state (on subnational resource transfers from the Peruvian Ministry of Economy and Finance, and conflict information provided by the National Ombudsman’s Office). From a substantive point of view, it provided the first evidence that “social” conflicts also had a distributive dimension. I do not think it is an exaggeration to say that this finding turned the study of socio-environmental conflicts on its head, challenging the emphasis on livelihood-based motivations for collective action that had been dominant in the Latin American literature [16–18]. Arellano-Yanguas’ paper was an early example of the methodological, empirical, and conceptual innovation that was possible using newly available bigger data.

The quantitative analysis of socio-environmental conflicts remains relatively limited in Latin America, with key contributions from Arce (2014), Arce et al. (2018), Arce, Polizzi, and Reeder (2020), Akchurin (2020), Arellano-Yanguas (2011), Haslam and Ary Tanimoune (2016), Haslam et al., (2018, 2019), Orihuela et al. (2018), Orihuela and Gamarra Echenique (2019), Ponce and McClintock (2014), Scheidel et al. (2020), and Sexton (2020) [19–29]. Taken as a group, these researchers have pushed the quantitative study of the resource curse to the subnational and local levels, and have done so by successfully integrating diverse sources of data including company characteristics; land-use patterns; surveys conducted by third parties; government statistics (censuses, financial transfers); coding of government reports such as environmental impact assessment documents; characteristics of the party system; and protest data available from the media, NGOs, academic projects, and, in the case of Peru, the National Ombudsman’s Office. For example, Akchurin [22] innovates by building a dataset of mining projects out of publicly available environmental impact assessment filings in Chile and crossing it with municipal data also made available by the state. By doing so, Akchurin was able to provide strong evidence for the strength of associational life in communities as a factor contributing to the likelihood of social mobilization. More detailed consideration of how these papers integrated bigger data, and particularly geographic information systems (GIS) and protest event analysis (PEA), is provided below.
3.1. Integration of Geographic Information Systems (GIS)

Geographic information systems (GIS) data has recently emerged as a powerful source of information. However, its analytical value for the study of social conflicts is principally in its combination with other economic, social, demographic, and institutional variables. In its bigger data version, characteristics of Earth’s surface are mapped continuously via satellite sensing, and are publicly available in datasets available from various sources including NASA and the European Space Agency (ESA). Increasingly, governments are making geo-referenced polygons available for extractive industry concessions, protected areas, indigenous territories, and municipal boundaries. Surface features can also be geo-referenced by hand-held devices and recorded, thus, allowing for technological updating of traditional mapping or participatory methods. Excellent reviews of the use of GIS techniques in the study of mining can be found in Werner et al. (2019) and Gleditsch and Weidmann (2012) for its application outside of geography and the Latin American region [30,31].

As evident in Werner et al.’s overview [30], GIS techniques have principally been used in the geographical sciences to identify some of the environmental impacts of mining, and to generate maps for illustrative purposes. On the first point, satellite imagery has been used to identify some environmental impacts of mining, notably coal mine fires [32], infrastructure development and deforestation [33–35], biodiversity loss [36], downstream water pollution [37], and risks associated with tailings dam placement. Such analyses are often restricted to particular mines and in this regard remain “small” [30] (p. 15).

More generally, GIS mapping has been used for visualization and communication of the impacts of extractive industry by both academics and civil society organizations [30,38]. Examples that would be familiar to many researchers working on socio-environmental conflicts include the interactive maps created by the Observatorio Latinoamericano de Conflictos Ambientales (OLCA) and the Environmental Justice (EJ) Atlas. Such mapping exercises can also be extremely sophisticated and technically demanding, as in the Oxfam-sponsored project to examine the overlapping geographies of mineral and oil and gas concessions with protected areas and indigenous lands in Peru and Ghana [39]. That project was able to identify the overlap between concessions and watershed drainage patterns that affected major river basins, successfully demonstrating “potential competition” between extractive and agricultural land uses [39] (p. 259).

However, we must also recognize that visualization alone offers limited purchase on understanding the social aspects of interest to qualitative researchers on social conflicts. For many social scientists, the key issues are related to the “linkages between human and natural systems” [38] (p. 1). Where GIS techniques have been used to link the geo-physical to the human, it is often at a small scale that basically complements traditional case study techniques. Such efforts may be viewed as a modern version of some classic mapping techniques long used by qualitative researchers, and may integrate participative techniques such as transect walks, interviews, and workshops [40–42]. For example, field sampling may also be integrated with topographic features and remote sensing to better understand how pollutants are likely to flow in groundwater [30]. Similarly, Patel et al. identify overlaps of large and small-scale mining using GIS, then qualitatively link those to academic, media, and civil society accounts of tensions between these sectors at a regional scale in Ghana [43].

These small-scale studies show that the key issue in linking GIS to the human is determining how to use geo-referencing as a peg that nails together different sources of data. Despite the important methodological innovations and substantive insights of these studies, the continuing focus on the local or subnational scale means that they tend to suffer from the same problems of case studies cited above. Put bluntly, despite the vast potential of GIS for the study of social conflicts, the leap from descriptive and illustrative richness to a rigorous assessment of causality has been much more difficult. The main focus of GIS techniques in the subfield, namely identifying overlaps or changes in land-use patterns over time, cannot be identified as generally causal to a class of social outcomes without the use of statistical methods and the integration of appropriate control variables.
In this regard, the main problem for the analysis of complex socio-environmental conflicts is the integration of different sources of data (physical and human) that is pegged to a place (geo-referenced). Socio-economic data is not yet generally available as GIS spatial layers [42] (p. 529), although we may expect this to change as national censuses are converted into geo-spatial datasets. An example of how GIS data can be pegged to other sources of human data is found in work by Haslam and Tanimoune (and Razeq). By geo-referencing the location of mining properties in five countries of Latin America, these researchers were able to combine localized GIS data on land use with census information at the municipal level, protest event data drawn from activist websites, and company-level information from corporate sources, using the mining property as a common locational peg [23–25]. The integration of land-use data from GIS sources permitted an objective assessment of the validity of the livelihoods thesis common in the literature, and the authors were able to offer a more nuanced and generalizable interpretation of the relationship between agricultural opportunities and social conflict.

Another innovative use of GIS pegs is found in Arce, Polizzi, and Reeder’s (2020) analysis of willingness to protest resource extraction, in which they cross the Latin American Public Opinion Project (LAPOP) survey’s respondent geo-reference with their proximity to geo-referenced mining projects [21]. This clever technique uses available data to substitute for running their own (very expensive!) survey of mining impact areas. The use of geo-referenced survey data allows them to make a much more compelling argument that (somewhat counter-intuitively) governance quality increases the willingness of individuals to protest against mining. Similar approaches are found in work on the social and environmental determinants of health, where researchers combine national health surveys with geo-referenced respondents (usually by municipality), and other sources of data [44], (p. 590). For example, proximity of a respondent to a mine can be used as a proxy measure of possible exposure to pollutants in the absence of sampling data [45] (p. 142). While random sampling for pollutants in the environment would be clearly preferred, it is often not feasible for the researcher [46,47].

Ponce and McClintock use survey data (also from the LAPOP survey) pegged at the regional level in Peru for their quantitative study [28] (pp. 128–129), which allows them to assess how individual satisfaction with services offered by the regional government correlated with participation in protest. In this regard, subnational data is essential to their overall argument that weak state capacity in Peruvian departments is associated with dissatisfaction and protest against extractive industries. Sexton [29] also argues that institutions matter to protest outcomes by suggesting that when strong they may mitigate community protest about environmental externalities. Part of his methodology uses geo-referenced mine site information available from the Peruvian government to categorize Peruvian provinces as either severely or less severely polluted.

These studies show that given some kind of geo-referential peg, there are vast opportunities to combine existing and publicly available data to conduct innovative and rigorous analyses of the intersection between mining projects (or any clearly delineated extractive project) and the natural and human environment.

3.2. Integration of Protest Event Analysis (PEA)

Protest event analysis (PEA) has emerged as a powerful tool for understanding socio-environmental conflicts related to resource extraction. PEA is a quantitative methodology theoretically grounded in the contentious politics literature [48–50]. PEA records “contentious events” that are reported in the media, and the unit of analysis is usually a specific act of protest on a certain date [50–53]. Reliance on media sources has traditionally meant the print media, but recent applications of the methodology have expanded the range of sources to include print, radio, television, and social media. In its more rigorous versions, each PEA event will identify claimants, target, type of contentious act, duration, and the number of participants, as in Arce [19]. However, the difficulty of capturing contentious events in the mining and resources sector, which are often rural and, therefore, under-reported, has led most scholars to rely on activist or governmental reports of conflict that cannot be disaggregated to an event-by-event timeline [14,15,23,28]. Arguably, the latter approach could be considered equivalent.
to coding campaigns [52] (p. 237). Be it “events” or “campaigns”, it is clear that most quantitative studies of socio-environmental conflicts require the use of an indicator of protest that must be coded by researchers from media or civil society sources.

Print media in Latin America has been historically capital-city centric, and there is good evidence from the United States and Europe that traditional print media under-reports contentious events, perhaps as few as ten percent of them [52,54]. For this reason, it makes sense to use alternative sources of information, particularly those provided by civil society organizations (CSOs). It is worth noting, however, that reliance on CSOs, especially national CSOs that aggregate data from affiliated organizations and individuals, may have a tendency to over-report protest. For example, it is well known that even when traditional print media and civil society organizations report on the same event, it is often characterized differently, even along apparently neutral indicators like the number of participants. It is relatively straightforward to search a media source or media aggregating platform (such as Factiva) for protest events during a given period, although this method will invariably give limited coverage of actual protest. Obtaining an accurate list of all protest events is almost impossible, and research with that objective in mind can involve a lot of fine-grained “detective work” drawing on multiple online sources. Reading and coding each news item into a PEA database is also tedious and time-consuming work. Such a database, once constructed, however, is a potential gold mine of information offering the potential to engage with wide-ranging research questions. Any researcher who has built these kinds of databases will tell you that it opens new research questions that had not been previously conceived.

Moreover, PEA data provides a standardized measure of how much protest is going on, and despite media and self-reporting (of activist organizations) biases, is potentially a good corrective to more subjective interpretations of activism based on case studies. Interest has been growing in the use of PEA to study social conflicts in Latin America, and there are both individual researchers and dedicated teams in almost every country that have developed and maintain private datasets with a focus on particular countries. These datasets are not formally consistent since there are methodological variations in how the data are collected, but common reference to the standard practices of contentious politics suggests that many categories of data will be similar, and may under certain circumstances be inter-operable. However, we should recognize that databases built around capital-city media with the original purpose of studying protest events from organized labor, or protests against neoliberalism, may provide limited information on rural protests against extractive industry.

Important contributions have been made to the study of social conflict using PEA; indeed, some measure of PEA is now the standard for identifying the occurrence of social conflict. As previously mentioned, Haslam and Ary Tanimoune [23] integrated PEA with an independently determined list of mining properties, and were, thus, able to include both conflict and non-conflict cases in their analysis, improving causal inference. In Peru, researchers have tended to rely on conflict accounts provided by the National Ombudsman’s Office [15,27–29]. Others have relied on activist sites that aggregate information about resource conflicts [14,20,23]. A notable exception is Arce, who used media reports to build his PEA database [19]. In this analysis, he was able to contextualize resource conflicts within a broader set of contentious politics, and link them to the opening of political opportunities [19].

Pan-regional databases on protest events that are prepped for quantitative analysis are extremely limited, and the only available source is the Social Conflict Analysis Database (SCAD) hosted at the Strauss Centre, University of Texas at Austin [55]. However, the Latin American version of the SCAD only covers Mexico, Central America, and the Caribbean. The EJ Atlas also maintains a database of conflict events that aggregates data from the major Latin American civil society organizations working on this topic—however, the data are provided publicly only as “visualization” and written case histories [14]. The “conflict” databases developed for the study of violent civil conflict, such as the Uppsala Conflict Data Program, are generally not considered appropriate to the study of social conflict in Latin America.
Another form of bigger data with potential utility for the analysis of socio-environmental conflicts is social media feeds such as those available from Facebook or Twitter—although this has attracted very little attention from researchers. Such sources may be considered an alternative measure of “protest events”, insofar as Twitter attacks and support tweets have been a key component of contemporary mobilization. Emerging methods of data mining social media permit the collection of data that can provide important information about the development of social networks [56,57]. Such techniques have not been widely applied to the study of resource conflicts despite ongoing interest in activist networks. However, the utility of these approaches is demonstrated by Varela’s (2019) study of the use (tweeting and retweeting) of Twitter hashtags originating with the anti-Agua Zarca dam movement in Honduras [58]. Varela used tweets to identify the international support networks for this movement, and analyze how the framing of their support differed from locally based organizations.

PEA datasets, whether organized around protest events or campaigns/conflicts, offer immense potential to understand mining and resource conflicts in the region. To date, most analyses have used conflict indicators coded from activist and civil society websites, which tend not to provide the fine-grained event-by-event descriptions traditionally available from media sources, but which offer insight into conflicts that would otherwise not be caught in a media search. Without a doubt, a pending task is coordination among the producers of existing conflict datasets and making them public in whole or in part—a task that is rendered more challenging by the intense amount of work that goes into producing one, and their great value once built. In the absence of such coordination and public access, researchers who want to use these databases will have to either negotiate access and appropriate acknowledgment with each PEA team, or build their own datasets from media and civil society sources.

4. Conclusions

We should also recognize that the collection and use of bigger data also raise ethical issues that can be distinct from fieldwork-based research. The most important ethical aspect of research is respect for human dignity, which may be disaggregated to include seeking “free, informed, and ongoing consent”; concern for the welfare of participants and their communities; and fair and equitable treatment of all people, especially those who are historically and structurally disadvantaged [59] (pp. 6–8). Typically, field researchers have been able to engage directly with participants to assess and achieve informed consent. Another concern of many field researchers relates to the question of local engagement and how we engage with research that considers local needs, respects their knowledge and traditions, and ultimately creates something of value for the people who are the “objects” of our research [60–63]. Integrating the use of bigger data with such ethical concerns may be quite difficult in practice.

With regards to the use of big data, the obvious ethical concerns are related to user-generated data that are captured by private companies without adequate consent, such as movement and call-log data from mobile phones [3] (p. 5). It is absolutely necessary that such data be anonymized, but also that researchers justify the use of such information in terms of contributing to a public “good”. In the case of bigger data that is already publicly available via postings on activist websites and social media, and reported in traditional media, we may certainly consider that participants have consented to such information being in the public domain.

Nonetheless, as we capture and transform this data (like coding PEA from textual reports by media, government institutions, or civil society organizations) we also transform activists’ own accounts of their activities into stories that are told to them and about them that may contribute to an “othering” with which some researchers may be uncomfortable. Raw data generated in Latin America by regular people and made public for the benefit of Latin Americans end up being captured by Northern-based academics, and processed to add value through analysis and publication in top-tier English-language journals [3] (p. 5). Is extracting data from individuals, NGOs, and public institutions regarding the nature and consequences of resource extraction another form of “extractivism” that maintains a relationship of exploitation and dependency? Such questions dovetail with ethical concerns about the
nature of the relationship between researcher and subject—although even the most self-consciously participative research cannot avoid being extractive at some level [64,65]. There may not be a right answer to these questions, but research using bigger data should also seek to engage with them and not simply assume there are no ethical issues to address.

This article aims to make the case that the new sources of bigger data that are available constitute an opportunity for researchers to better understand the complex problem of socio-environmental conflict. In this article, I underlined that the key problem in the exploitation of bigger data for the analysis of socio-environmental conflict is how to mesh different sources of data together. The localized nature of mining and other resource conflicts makes the use of geo-referential pegs across diverse sets of data a logical way to organize diverse sources of data. The variety of this bigger data must be underlined, and include surveys, remote (GIS) sensing, social media accounts, structured datasets from organizations and governments, and textual data that need to be interpreted and coded (from media accounts of conflict, to activist websites, and environmental impact assessments). Methodologically, the use of bigger data is an opportunity to solve problems of selection and subjective bias created by the selection of cases on the dependent variable, and the selection of cases with extreme values. Triangulation of data with the aim of producing insights that are valid and generalizable with a high degree of confidence should be a central concern of all researchers. The new world of bigger data is, thus, a wonderful opportunity to explore new dimensions of socio-environmental conflict, in which the sources and uses of this data are only limited by our imaginations.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

**References**

1. Günther, W.A.; Rezazade Mehrizi, M.H.; Huysman, M.; Feldberg, F. Debating big data: A literature review on realizing value from big data. *J. Strateg. Inf. Syst.* 2017, 26, 191–209. [CrossRef]
2. Hilbert, M. Big data for development: A review of promises and challenges. *Dev. Policy Rev.* 2016, 34, 135–174. [CrossRef]
3. Mann, L. Left to other peoples’ devices? A political economy perspective on the big data revolution in development. *Dev. Chang.* 2017, 49, 3–36. [CrossRef]
4. George, A.; Bennett, A. Case Studies and Theory Development in the Social Sciences; MIT Press: Cambridge, MA, USA, 2005.
5. King, G.; Keohane, R.; Verba, S. *Designing Social Inquiry: Scientific Inference in Qualitative Research*; Princeton University Press: Princeton, NJ, USA, 1994.
6. Mahoney, J. After KKV: The new methodology of qualitative research. *World Politics* 2010, 62, 120–147. [CrossRef]
7. Lijphart, A. Comparative politics and the comparative method. *Am. Political Sci. Rev.* 1971, 65, 682–693. [CrossRef]
8. Lieberson, S. Small N’s and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases. *Soc. Forces* 1991, 70, 307–320. [CrossRef]
9. Sartori, G. Concept misinformation in comparative politics. *Am. Political Sci. Rev.* 1970, 64, 1033–1053. [CrossRef]
10. Amengual, M. Buying Stability: The distributive outcomes of firm responses to risk in the Bolivian mining industry. *World Dev.* 2018, 104, 31–45. [CrossRef]
11. Steinberg, J. Mining, Communities, and States: The Local Politics of Natural Resource Extraction in Africa; Cambridge University Press: Cambridge, UK, 2019.
12. CCSRC (Canadian Centre for the Study of Resource Conflict). *Corporate Social Responsibility: Movements and Footprints of Canadian Mining and Exploration Firms in the Developing World*; CCSRC: Revelstoke, BC, Canada, 2009.
13. Imai, S. The ‘Canada brand’: Violence and Canadian mining companies in Latin America. *Osgoode Leg. Stud. Res. Pap.* 2017, 13, 129. [CrossRef]
14. Scheidel, A.; Del Bene, D.; Liu, J.; Navas, G.; Mingorria, S.; Demaria, F.; Avila, S.; Roy, B.; Erkütör, I.; Temper, L.; et al. Environmental conflicts and defenders: A global overview. *Glob. Environ. Chang.* 2020, 63, 102–104. [CrossRef]

15. Arellano-Yanguas, J. Aggravating the resource curse: Decentralization, mining and conflict in Peru. *J. Dev. Stud.* 2011, 47, 617–638. [CrossRef]

16. Bebbington, A.; Humphreys Bebbington, D.; Bury, J.; Lingan, J.; Muñoz, J.P.; Scurrah, M. Mining and social movements: Struggles over livelihood and rural territorial development in the Andes. *World Dev.* 2008, 36, 2888–2905. [CrossRef]

17. Bridge, G. Mapping the bonanza: Geographies of mining investment in an era of neoliberal reform. *Prof. Geogr.* 2004, 56, 406–421.

18. Bury, J. Livelihoods, mining and peasant protests in the Peruvian Andes. *J. Lat. Am. Geogr.* 2002, 1, 3–18. [CrossRef]

19. Arce, M. *Resource Extraction and Protest in Peru*; University of Pittsburgh Press: Pittsburgh, PA, USA, 2014.

20. Haslam, P.; Ary Tanimoune, N. The determinants of social conflict in the Latin American mining sector: New evidence with quantitative data. *World Dev.* 2016, 79, 401–419. [CrossRef]

21. Haslam, P.A.; Ary Tanimoune, N.; Razeq, Z. Is ‘being foreign’ a liability for mining companies? locational liabilities and social conflict in Latin America. *Resour. Policy* 2019, 63, 101425. [CrossRef]

22. Haslam, P.A.; Ary Tanimoune, N.; Razeq, Z. Do Canadian mining firms behave worse than other companies? Quantitative evidence from Latin America. *Can. J. Political Sci.* 2018, 51, 521–551.

23. Orihuela, J.C.; Gamarra Echenique, V. Volatile and spatially varied: The geographically differentiated economic outcomes of resource-based development in Peru, 2001–2015. *Extr. Ind. Soc.* 2019, 6, 1143–1155. [CrossRef]

24. Orihuela, J.C.; Pérez, C.A.; Huaroto, C. Do fiscal windfalls increase mining conflicts? Not always. *Extr. Ind. Soc.* 2018, 6, 313–318. [CrossRef]

25. Ponce, A.; McClintock, C. The explosive combination of inefficient local bureaucracies and mining production: Evidence from localized social protests in Peru. *Lat. Am. Politics Soc.* 2014, 56, 118–139. [CrossRef]

26. Sexton, R. Unpacking the local resource curse: How externalities and governance shape social conflict. *J. Confl. Resolut.* 2020, 64, 640–673. [CrossRef]

27. Werner, T.T.; Bebbington, A.; Gregory, G. Assessing impacts of mining: Recent contributions from GIS and remote sensing. *Extr. Ind. Soc.* 2019, 6, 993–1012. [CrossRef]

28. Gleditch, K.S.; Weidmann, N.B. Richardson in the information age: Geographic Information Systems and spatial data in international studies. *Annu. Rev. Political Sci.* 2012, 15, 461–481. [CrossRef]

29. Biswal, S.S.; Raval, S.; Gorai, A.K. Delineation and mapping of coal mine fire using remote sensing data—A review. *Int. J. Remote Sens.* 2019, 40, 6499–6529. [CrossRef]

30. Goparaju, L.; Prasad, P.R.C.; Ahmad, F. Geospatial technology perspectives for mining vis-à-vis sustainable forest ecosystems. *Present Environ. Sustain. Dev.* 2017, 11, 219–238. [CrossRef]

31. Suh, J.; Kim, S.-M.; Yi, H.; Choi, Y. An overview of GIS-based modeling and assessment of mining-induced hazards: Soil, water, and forest. *Int. J. Environ. Res. Public Health* 2017, 14, 1463. [CrossRef]

32. Bebbington, A.J.; Humphreys Bebbington, D.H.; Sauls, L.A.; Rogan, J.; Agraval, S.; Gamboa, C. Resource extraction and infrastructure threaten forest cover and community rights. *Proc. Natl. Acad. Sci. USA* 2018, 12505, 13164–13173. [CrossRef]

33. Murguia, D.L.; Bringeuz, S.; Schaldach, R. Global direct pressures on biodiversity by large-scale metal mining: Spatial distribution and implications for conservation. *J. Environ. Manag.* 2016, 180, 409–420. [CrossRef]

34. Alonzo, M.; Van Den Hoek, J.; Ahmed, N. Capturing coupled riparian and coastal disturbance from industrial mining using cloud-resilient satellite time series analysis. *Sci. Rep.* 2016, 6, 1–12. [CrossRef] [PubMed]

35. Lechner, A.M.; Owen, J.; Ang, M.; Kemp, D. Spatially integrated social sciences with qualitative GIS to support impact assessment in mining communities. *Resources* 2019, 8, 47. [CrossRef]
39. Cuba, N.; Bebbington, A.; Rogan, J.; Millones, M. Extractive industries, livelihoods and natural resource competition: Mapping overlapping claims in Peru and Ghana. *Appl. Geogr.* 2014, 54, 250–261. [CrossRef]
40. Bar-On, A.A.; Prinsen, G. Planning, communities and empowerment: An introduction to participatory rural appraisal. *Int. Soc. Work* 1999, 42, 277–294. [CrossRef]
41. Chambers, R. Participatory Rural Appraisal (PRA): Analysis of experience. *World Dev.* 1994, 22, 1253–1268. [CrossRef]
42. Scheueler, V.; Kuenmerle, T.; Schröder, H. Impacts of surface gold mining on land use systems in Western Ghana. *Ambio* 2011, 40, 528–539. [CrossRef] [PubMed]
43. Patel, K.; Rogan, J.; Cuba, N.; Bebbington, A. Evaluating conflict surrounding mineral extraction in Ghana: Assessing the spatial interactions of large and small-scale mining. *Extr. Ind. Soc.* 2016, 3, 450–463. [CrossRef]
44. Riva, M.; Terashima, M.; Curtis, S.; Shucksmith, J.; Carlebach, S. Coalfield health effects: Variation in health across former coalfield areas in England. *Health Place* 2011, 17, 588–597. [CrossRef]
45. Saha, S.; Pattanayak, S.K.; Sills, E.O.; Singha, A.K. Under-mining health: Environmental justice and mining in India. *Health Place* 2011, 17, 140–148. [CrossRef]
46. Taylor, M.P.; Mackay, A.K.; Hudson-Edwards, K.A.; Holz, E. Soil Cd, Cu, Pb and Zn contaminants around Mount Isa city, Queensland, Australia: Potential sources and risks to human health. *Appl. Geochem.* 2010, 25, 841–855. [CrossRef]
47. Wyatt, L.; Ortiz, E.J.; Feingold, B.; Berkly, A.; Diringer, S.; Morales, A.M.; Jurado, E.R.; Hsu-Kim, H.; Pan, W. Spatial, temporal, and dietary variables associated with elevated mercury exposure in Peruvian riverine communities upstream and downstream of artisanal and small-scale gold mining. *Int. J. Environ. Res. Public Health* 2017, 14, 1582. [CrossRef]
48. McAdam, D.; Tarrow, S.; Tilley, C. *Dynamics of Contention*; Cambridge University Press: Cambridge, UK, 2001.
49. Tarrow, S. *Power in Movement: Social Movements and Contentious Politics*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2008.
50. Tilly, C. *Contentious Performances*; Cambridge University Press: Cambridge, UK, 2008.
51. Beissinger, M.R. Nationalist violence and the state: Political authority and contentious repertoires in the former USSR. *Comp. Politics* 1998, 30, 401–422. [CrossRef]
52. Koopmans, R.; Rucht, D. Protest event analysis. *Mobilization* 1999, 4, 231–259.
53. López-Mayoa, M. Venezuela after the Caracazo: Forms of protest in a deinstitutionalized context. *Bull. Lat. Am. Res.* 2002, 21, 199–218. [CrossRef]
54. Earl, J.; Martin, A.; McCarthy, J.D.; Soule, S. The use of newspaper data in the study of collective action. *Annu. Rev. Sociol.* 2004, 30, 65–80. [CrossRef]
55. Salehyan, I.; Hendrix, C.S.; Hamner, J.; Case, C.; Linebarger, C.; Stull, E.; Williams, J. Social conflict in Africa: A new database. *Int. Interact.* 2012, 38, 503–511. [CrossRef]
56. Zafarani, R.; Abbasi, M.A.; Liu, H. *Social Media Mining: An Introduction*; Cambridge University Press: Cambridge, UK, 2014.
57. Szabo, G.; Polatkan, G.; Boykin, P.O. *Social Media Data Mining and Analytics*; John Wiley & Sons, Inc.: Indianapolis, IN, USA, 2018.
58. Varela, Y. Megadam Resistance and the Two Solitudes of Activism: Lenca Resistance to the Agua Zarca Megadam in Honduras. In *M.A. Major Research Paper*; School of International Development and Global Studies, University of Ottawa: Ottawa, ON, Canada, 2019.
59. Tri-Council. *Ethical Conduct for Research Involving Humans*. In *Policy Statement*; Government of Canada: Ottawa, ON, Canada, 2014. Available online: https://ethics.gc.ca/eng/documents/TCPS_2-2014_FINAL_Web.pdf (accessed on 16 September 2020).
60. Harrison, M.E. Collecting sensitive and contentious information. In *Doing Development Research*; Desai, V., Potter, R., Eds.; Sage: Thousand Oaks, CA, USA, 2006; pp. 62–69.
61. Banks, G.; Scheyvens, R. Ethical issues. In *Development Fieldwork*, 2nd ed.; Scheyvens, R., Ed.; Sage Publications: Thousand Oaks, CA, USA, 2014.
62. Momsen, J.H. Women, men and fieldwork: Gender relations and power structures. In *Doing Development Research*; Desai, V., Potter, R., Eds.; Sage: Thousand Oaks, CA, USA, 2006; pp. 44–51.
63. Schafer, J.; Haslam, P.A.; Beaudet, B. Meaning, measurement, and morality in international development. In *Introduction to International Development: Approaches, Actors, Issues, and Practice*, 3rd ed.; Haslam, P., Schafer, J., Beaudet, P., Eds.; Oxford: Don Mills, ON, Canada, 2017; pp. 2–24.
64. Cornwall, A.; Pratt, G. The use and abuse of participatory rural appraisal: Reflections from practice. *Agric. Hum. Values* **2011**, *28*, 263–272. [CrossRef]

65. Leurs, R. Critical reflections on Rapid and Participatory Rural Appraisal. *Dev. Pract.* **1997**, *7*, 290–293.