Chapter Four: The Good, the Bad and the Beauty of ‘Good Enough Data’

Miren Gutiérrez

# Abstract

Drawing on the concept of ‘good enough data,’[[1]](#footnote-1) which apply to citizen data collected via sensors this chapter looks critically at data in ‘proactive data activism,’ understood as a social practice that uses the data infrastructure politically and proactively to foster social change.[[2]](#footnote-2) This chapter examines how data are generated and employed in proactive data activism, expanding and applying the term ‘good enough data’ beyond citizen sensing and the environment. This analysis derives from a taxonomy of activists based on how they go about obtaining data.[[3]](#footnote-3) It offers too an unsentimental view on the failures and contradictions of data activism regarding the collection, analysis and communication of data.[[4]](#footnote-4) The chapter employs the Syrian Archive –an organization that curates and documents data related to the Syrian conflict for activism— as a pivotal case to look at the new standards applied to data gathering and verification in data activism from the South, as well as their challenges, so data become ‘good enough’ to produce reliable evidence for social change. Data were obtained too thorough in-depth interviews, fieldwork and empirical observation.

# Introduction

On 25 and 30 March 2017, the town of al-Lataminah, in northern Syria, suffered two chemical attacks possibly committed by Syrian government aircraft.[[5]](#footnote-5) Since 2011, Syria has been engulfed by a multi-sided armed conflict between forces led by President Bashar al-Assad, together with its international allies, and other forces opposing both the government and each other in varying groupings.[[6]](#footnote-6) International organizations have blamed the Syrian government, and its ally, Russia, opponent rebel groups and the coalition led by the United States of human rights violations,[[7]](#footnote-7) including attacks with chemical weapons against civilians. The main forces implicated in the Syrian conflict –that is, Iran, Russia, Syria and the United States— are parties to the Weapons Convention, which entered into force in 1997, prohibiting the production, possession and use of chemical weapons, and should comply with its obligations.[[8]](#footnote-8) However, since 2013, there have been some forty recorded instances of alleged chemical weapons use in Syria,[[9]](#footnote-9) and no mechanism to attribute responsibility exists.[[10]](#footnote-10)

As with other assaults, the source for the attacks against al-Lataminah remained unclear until a team of students at the Berkeley Human Rights Investigation Lab (HRC Lab) went through nine videos uploaded by people identified as journalists and ordinary citizens on Twitter, YouTube and Facebook. This analysis suggested that the attacks had indeed involved chemical weapons and that, at least in the case of the 25 March strike, the target was a medical facility.[[11]](#footnote-11) Their report concludes that the perpetrators were in potential violation of international humanitarian law, as well as other regulations and treaties.[[12]](#footnote-12)

*Figure 1: Chemical Strikes on Al-Lataminah. Source: (Syrian Archive, 2018).*

Syrian Archive was the supplier of the videos. This organization gathers visual accounts about human rights violations committed by all sides of the Syrian conflict.[[13]](#footnote-13) It was founded in 2014 by human rights activists with the aim of investigating and preserving digital content as a form of ‘a digital memory’ of human rights infringements and building a corpus of substantiation that could be used for advocacy, reporting and law cases.[[14]](#footnote-14) Syrian Archive is sustained by donations but accepts no money from governments directly involved in the conflict.[[15]](#footnote-15)

Syrian Archive could be regarded as a proactive data activist organization, or activism that employs the data infrastructure politically and proactively to provide diagnoses and evidence for social change.[[16]](#footnote-16) The massive collection and automated analysis of private data by governments and corporations have generated the emergence of reactive data activists, who use the data infrastructure to shelter their interactions from data surveillance.[[17]](#footnote-17) The chapter is focused on the data practices employed by the former type of activism, which embodies a reversal to what Poell, Kennedy and van Dijck call *dataveillance.*[[18]](#footnote-18) In the face of discriminatory, opaque or unethical private data gathered for obscure purposes without people’s knowledge or consent, proactive data activists engage in a variety of methods to obtain and analyze data, sidestepping conventional data actors and systems, enhancing their data agency, and correcting the asymmetries embedded in top-down datafication processes.[[19]](#footnote-19)

Activists have demonstrated ingenuity in creating data, and can be classified from the ways they go about obtaining them: from the easiest way to the most difficult, they can rely on whistleblowers who hand over datasets anonymously (first type) or resort to public datasets (second); and when data are not available, create platforms to crowdsource citizen data (third); appropriate data, becoming whistleblowers themselves (fourth), and generate data on their own terms, for example, via data-gathering devices (fifth).[[20]](#footnote-20)

Syrian Archive combines mainly the second and third kinds of data extraction methods: it counts on data uploaded by ordinary citizens, journalists and activists on sharing and social media platforms. That is, it relies on data that has been made public by people (i.e. a form of public data). These data have been harvested through platforms which were not set up by the activists themselves but by social media service providers (i.e. a form of crowdsourced data). Other data activists specialize in using just one of these methods.[[21]](#footnote-21) For example, the deployers of ‘Ayuda Ecuador’ –an application of the Ushahidi platform launched to gather data on the crisis unleashed by the earthquake in 2016 in Ecuador— established the means to collect data submitted via mobile technology, emails and websites, using the third type of data extraction method. Although crowdsourcing citizen data is its main data mining method, Ushahidi also resorts to data scraping from websites and social media for verification purposes (i.e. a form of data appropriation). Meanwhile, the citizen data on air quality Gabrys, Pritchard and Barratt talk about were gathered via sensors,[[22]](#footnote-22) namely, the fifth technique. Data activists are resourceful and often combine repertoires of action and strategies seeking the best solutions for each case.

Proactive data activism often relies on good enough data, a concept coined by Gabrys, Pritchard and Barratt to indicate data created, valued and interpreted by ordinary people, which these authors apply to environmental data and citizen sensing.[[23]](#footnote-23) The good enough data practices in these authors’ study meet the criteria of being a) based on datasets gathered by ordinary people, b) aimed at creating new types of data and data stories, and c) useful to generate evidence that can be employed in decision-making.[[24]](#footnote-24) This concept pertains to other types of data activism too, as seen later.

However, data activism is neither pure nor neutral; it embeds some contradictions and unbalances. For example, like in other deployments of the Ushahidi platform, ‘Ayuda Ecuador’ places digital humanitarians (i.e. the deployers of the application) in the role of gatekeepers of good enough data as they are the ones enforcing the data authentication systems and controlling the crisis map. These data asymmetries are reviewed later too.

To look at good enough data in activism, I rely on fieldwork, empirical observation of significant cases, a case study, and in-depth interviews with relevant actors involved in data curation processes from several organizations. The case study methodology aims at showing how data practices look like based not on the actors’ opinions, but on evaluations, scientific literature and website content.[[25]](#footnote-25) Syrian Archive is a data activist initiative that embodies the highest standards of data curation in activism. Because of the delicate tasks they have been set to confront, its managers have developed a meticulous data protocol, based on international standards, to conserve data so they are *good enough* to sustain further legal investigations. This is a single-focused, descriptive study that employs an instance to offer a common framework for future studies.[[26]](#footnote-26) Syrian Archive provides a case of a class of phenomena; that is, activism that utilizes the data infrastructure for knowledge and action. The case study methodology is helpful when the boundaries between a phenomenon and its context are not apparent.[[27]](#footnote-27) This organization has been chosen based on the fact that background conditions are relevant to the inspection of data practices in activism.

Five interviewees –chosen for their in-depth knowledge data practices and protocols in activism and beyond— were questioned about their data mining and curation methodologies. Although the questions vary in line with their expertise, the questionnaires sent to these practitioners include open questions such as: ‘How can you aspire to data robustness taking into account that your data sources may have their own agendas?’ The idea behind the interviews is to capture what these interviewees think they do.[[28]](#footnote-28) The questions were designed to last for one hour, and sent by email or formulated in phone discussions. Interviewee 1 works for a crowdsourcing platform that visualizes citizen data for digital humanitarianism. Interviewee 2 has designed smart grids that channel data for real-time decision-making in water management. Interviewees 3 and 4 have an exhaustive knowledge of Syrian Archive. In charge of an index that ranks countries by their frailty, Interviewee 5 –not a data activist— has been included here for contrast, as in the case of Interviewee 2. Although their data methods and approaches vary, their tenets do not differ so much.

The next sections offer an exploration of data in activism from the perspective of data, followed by an inspection of Syrian Archive and a comparison with other initiatives, concluding with a proposed definition for *good enough data in activism* that can be used as a heuristic tool in other analyses.

## Perfect Data, Big Data and Small Data

Most researchers, journalists and activists who handle data employ ‘small data’, namely, data that appears in a volume and format that makes them usable and analyzable.[[29]](#footnote-29) In contrast, ‘big data’ are so vast and complex that they cannot be managed with traditional data processing methods.[[30]](#footnote-30) The big data infrastructure allows the collection of data and metadata continually, surreptitiously and comprehensively: every click and every ‘like’ is stored and analyzed in real-time.[[31]](#footnote-31) Big data are understood as ‘datasets whose volume, velocity or variety’ is extremely high and show a high degree of ‘veracity’ and a potential for ‘value’.[[32]](#footnote-32)

The employment of big data techniques can give rise to thought-provoking insights. For example, Stephens-Davidowitz’s *Everybody Lies* explores how people lie in polls and pretend on social media while removing their masks when they search for information online.[[33]](#footnote-33) Based on the massive analysis of Google searches, other search engines and websites, he discovers that people are much more racist, sexist and ignoble than they think or admit.[[34]](#footnote-34) Big data-based research is revolutionizing not only surveillance and marketing, but scientific disciplines as well, from sociology to historiography. For instance, only a few years ago it was only possible to examine fragments of big *corpora* and science focussed on getting information from limited fragments.[[35]](#footnote-35) The study of the 13th century Korean Buddhist canon, which contains 52 million characters distributed across 166,000 pages, is an example; with data analytics, the whole text can now be interrogated in its totality in every search.[[36]](#footnote-36) Are big data, then, better data, as Cukier argues?[[37]](#footnote-37) The simple answer is: not necessarily. Data *goodness* depends on the quality of both the data and processes involved to analyze them, as well as on the purpose of the analysis. Namely, it is not a question only of quantity, but also of quality and intent.

No matter how big, data can never be perfect because they cannot be raw.[[38]](#footnote-38) Borges’s tale of frustrated mapmakers could be read as a warning about the impossibility of perfect science, or data.[[39]](#footnote-39) In ‘On Exactitude in Science,’Borges speaks about an empire where pursuing scientific authenticity, cartographers made maps as big as the land they portrayed, whereupon these charts failed to be useful and were forgotten.[[40]](#footnote-40) The same way that ‘the map is not the territory’,[[41]](#footnote-41) data cannot be considered ‘straightforward representations of given phenomena’.[[42]](#footnote-42) Data are not free of ideology, as they do not emerge free of the views, methods and technologies of the people that conceive, generate, process, curate, analyze and store them.[[43]](#footnote-43) Data are not natural but *cultural resources* ‘cooked’ in processes of collection and use, which are also ‘cooked’,[[44]](#footnote-44) thus, they can embed biases, gaps and asymmetries. As Interviewee 5 notes, ‘every qualitative data source is arguably open to bias.’ Paraphrasing Gitelman,[[45]](#footnote-45) *perfect data* is another oxymoron.

Errors and biases can be introduced accidentally in data processes. An artificial intelligence (AI) algorithm learned to associate women with kitchen images because more women appear photographed in kitchens on the web.[[46]](#footnote-46) In the process of *learning,* the algorithm then multiplied the bias present in the initial dataset, amplifying –not simply replicating— the biased association between kitchens and women.[[47]](#footnote-47) This study by the University of Virginia is among several others that recently show that artificial intelligence systems can incorporate, and even multiply, biases if their design or the data on which they are based are not revised carefully.

Besides, data can also be manipulated. Edward Snowden’s 2013 revelations include examples of big data-based manipulation; they showed that the intelligence services of the US government, with the collaboration of companies and other governments, had established a dense layer of surveillance and data interception on the communications of millions of people globally without their knowledge.[[48]](#footnote-48) More recently, it was revealed that Cambridge Analytica, a data analytics firm, had collaborated with Donald Trump’s election team in 2016 to harvest millions of Facebook profiles of US voters to design personalized campaign messages for them.[[49]](#footnote-49) Critical scholars, including Braman,[[50]](#footnote-50) Tufekci,[[51]](#footnote-51) and Gangadharan,[[52]](#footnote-52) describe how big data techniques are employed to profile people, discriminate against vulnerable groups and promote relentless, omnipresent and preventive monitoring. In *Weapons of Math Destruction*, O'Neil presents too that data-based programs increase the efficiency of ‘predatory advertising’ that undermines democracy.[[53]](#footnote-53)

This paper’s aim is not to determine the difference between *good* and *good enough* because to do that, and it would have to establish first the philosophical question of the meaning *good,* over which there is no consensus*.* Gabrys, Pritchard and Barrat do not fathom what *good* means related to data either.[[54]](#footnote-54) Thus, good enough is to be initially understood asrobust enough - devoid of significant unfair or impairing manipulation - to sustain research for socially beneficial purposes. Coming back to Cukier’s assertion,[[55]](#footnote-55) it seems that instead of asking how big our datasets are to determine whether they are good, we should start by asking first ‘for what’, as Robert Staughton Lynd suggested about science,[[56]](#footnote-56) and then look for data and processes that are strong enough to support such research.

## Good Enough Data

Data analysis is not for everyone. In a show of enthusiasm, Rogers compared data analysis with punk music in 2012 because ‘anyone can do it’.[[57]](#footnote-57) Unfortunately, this is not the case. Not only are data not available to anyone, but, as Innerarity points out ,[[58]](#footnote-58) the tools to make them useful are not available to anyone either. However, the barriers to access both data and their infrastructure –understood as the software, hardware and processes necessary to render them useful— have been falling in recent years.[[59]](#footnote-59) Activists of all stripes are using data, in combination with the geoweb, satellite technology, sensing devices and information and communication technologies (ICTs), to create datasets and generate diagnoses and social change, as I have considered elsewhere.[[60]](#footnote-60) Examples abound. Forensic Architecture, supported by Amnesty International, has created an interactive model of Syria's most notorious prison, Saydnaya, using the memories of sounds narrated by survivors who had been kept captive in the dark. The project aims to show the conditions inside the prison.[[61]](#footnote-61) WeRobotics employs ‘community drones’ to collect data on the conditions of glaciers in Nepal to analyze the data and launch alarms when there is the danger of flash floods.[[62]](#footnote-62) InfoAmazonia, among other things, has published a calendar that superimposes the time counted by the Indigenous peoples of the Tiquié River, Brazil, and the time measured in the Gregorian calendar, in a dialogue between the local and the global that never existed before.[[63]](#footnote-63) Meanwhile, the Ushahidi platform is deployed to generate crisis maps that can visualize citizen data in quasi-real-time to assist humanitarian operations. Namely, when data are not available, people and organizations create them.

Gabrys, Pritchard and Barratt refer to ‘just good enough’ data generated by non-experts to gauge pollution.[[64]](#footnote-64) Citizen data are often good enough to produce ‘patterns of evidence’ that can mobilize community responses to connect with regulators, request follow-up monitoring, make a case for improved regulation and accountability, and keep track of exposures both on an individual and collective level.[[65]](#footnote-65) Although these authors imply citizen sensing and environmental data gathering, the concept of good enough data can be applied to other areas where citizens and organizations are creating their own datasets, robust enough to generate new types of data and data practices, and impactful bottom-up narratives. The same way that alternative journalism argues that different forms of knowledge may be generated, representing ‘multiple versions of reality from those of the mass media’,[[66]](#footnote-66) good enough data – as an alternative to conventional big data - can produce trustworthy analysis and visualizations which shed light on complex issues and are the basis for action. The examples of Syrian Archive, Forensic Architecture, WeRobotics, InfoAmazonia and Ushahidi show that, beyond citizen sensing, activists can produce reliable results based on good enough data in the areas of human rights evidence, crisis mapping and alert systems, and advocacy. The selection of these cases - based on the complexity in their data processes - draws from my previous work on data activism.[[67]](#footnote-67)

That is not to say that data in activism is devoid of inconsistencies. Some Ushahidi deployments, for instance, have been set up and then abandoned,[[68]](#footnote-68) creating a cyber-graveyard of ‘dead Ushahidi’ maps for lack of communities reporting data.[[69]](#footnote-69) Also, Ushahidi’s verification system has not forestalled some glitches, with on-the-ground consequences.[[70]](#footnote-70) Besides, data activist projects often resort to proprietary, corporate inventions, with entrenched imbalances and harmful practices (i.e. semi-slave labor in the extraction of the cobalt needed in smartphones),[[71]](#footnote-71) which are then embedded in the initiative.[[72]](#footnote-72) Data processes in activism can also include asymmetries in the relationships established within their networks, integrating their own ‘power inequalities, unbalances and mediations’.[[73]](#footnote-73) The interviewees in this study admit that faulty datasets must be corrected. However, constant verification allows imperfect datasets to improve and lead to alternative and useful insights.

Summarizing, good enough data in activism could be explained as data, however flawed, which are obtained, gathered, curated and analyzed by citizens and activists through procedures aimed at generating action-oriented analysis for beneficial social change and humanitarianism. This definition is reviewed later.

## Syrian Archive

In 2016, Syrian Archive team verified 1748 videos and published a report pointing to an ‘overwhelming’ Russian participation in the bombardment and airstrikes against civilians in the city of Aleppo, Syria.[[74]](#footnote-74) Although all parties have perpetrated violations, the visual evidence demonstrated that the Russian forces were accountable for the largest amount of violations in Aleppo.[[75]](#footnote-75) Meanwhile, the Office of the UN High Commissioner for Human Rights (OHCHR) issued a carefully phrased statement in which it blamed ‘all parties to the Syrian conflict’ of perpetrating violations resulting in civilian casualties, although it also admitted that ‘government and pro-government forces’ (i.e. Russian) were attacking hospitals, schools and water stations.[[76]](#footnote-76) The difference in language could be attributed to the attitude of a non-governmental organization compared with that of a UN agency, which has to collaborate with governments to be able to function. However, the contrast in data methodologies may also be at the bottom of what each said about the bombings. While the OHCHR report was based on interviews with people after the events, Syrian Archive relied on video evidence mostly uploaded by people on social media without the intervention of the non-profit (although some video evidence was sent directly to Syrian Archive).[[77]](#footnote-77)

The traditional practices of evidence gathering in human rights law ‘is grounded in witness interviews often conducted well after the fact’.[[78]](#footnote-78) Even if surveys and interviews offer crucial information about armed conflicts, the violence can shape responses ‘in ways that limit their value’.[[79]](#footnote-79) Namely, the distortions that violence can elicit in the witnesses’ testimonies and the time-lapse between the facts and the interviews can affect the reliability of an inquest. The medialization of conflicts makes the new activist’s data gathering techniques relevant. Syrian Archive does not only resort to what victims and witnesses say they remember, but they also rely on what witnesses *recorded as events unfolded* in real-time. Most of the evidence included in Syrian Archive report on Aleppo was mentioned in the OHCHR report, but the non-profit also found new evidence that had not been cited before.[[80]](#footnote-80)

Syrian Archive’s methodology used in carrying out the Aleppo bombings research include the identification, collection and preservation of data, followed by two layers of confirmation with increasing depth.[[81]](#footnote-81) Syrian Archive has identified three hundred sources in Syria - including media organizations and citizen journalists - and checked their reliability by tracking them and examining their social media accounts over time.[[82]](#footnote-82) The organization is aware of the fact that these sources are ‘partisan’ and require ‘caution’, and that is why it also relies on other groups of sources, which offer additional information.[[83]](#footnote-83)

When Syrian Archive obtains a video, the first thing is to save it; next, the identity of the source of the video is verified with the on-the-ground network. Then, the source’s track record is examined looking at whether the source has posted other videos from the same location or different locations.[[84]](#footnote-84).As investigative reporters counting on deep throats, Syrian Archive does not scrutinize the sources’ motivations, as Interviewee 4 notes:

We don’t care about their agendas. What we are trying to do here is to look at the content itself, and cross-reference the content with hundreds of thousands of other contents to conclude whether it is true or not. We are aware of the types of problems that our sources endure (for example blockages by the Syrian government), but we do not look at that. As much as we can understand limitations for sources, our job is to verify the contents.

Subsequently, the recording is examined to establish its location using methods that can include the identification of a natural feature of the landscape (e.g. a line of trees) and buildings and landmarks (e.g. a mosque’s minaret).[[85]](#footnote-85) Syrian Archive compares the footage with satellite images from Google Earth, Microsoft Bing, OpenStreetMap, Panoramio or DigitalGlobe, and looks at the metadata from the video, which also can provide information about its whereabouts. The video is then compared with the testimonies of witnesses interviewed by trustworthy media outlets and human rights organizations. Videos in Arabic are also scrutinized to determine the dialect and location to which they might be linked. When possible, Syrian Archive contacts the source directly to confirm the location and teams up with organizations specialized in verifying images, such as Bellingcat.[[86]](#footnote-86) The materials are categorized using the classification of human rights abuses issued by the UN, which includes, among others, ‘massacres and other unlawful killings’, ‘arbitrary arrest and unlawful detention’ and ‘hostage-taking’.[[87]](#footnote-87) On Syrian Archive website, information is classified by the kind of weapon utilized, the location or the kind of violation or abuse. Figure 2 summarizes the process.

*Figure 2: Syrian Archive digital evidence workflow[[88]](#footnote-88).*

Based on the examination of footage, the research of the Aleppo bombings from July to December 2016 includes overwhelming evidence of ‘unlawful attacks,’ illegal weapons, attacks against civilians, humanitarian workers, journalists and civilian facilities such as hospitals, schools or markets, and the use of incendiary and cluster munitions.[[89]](#footnote-89) Given that rebel groups and terrorist militias like the Islamic State do not have air forces, Syrian Archive concludes that most of the airstrikes are carried out by Russian aircraft.[[90]](#footnote-90) Figure 3 shows a static snapshot of the location of verified attacks; the interactive chart available online allows users to access the videos that sustain it.

*Figure 3: Map of visual evidence about attacks in Aleppo City (July - December 2016).[[91]](#footnote-91)*

Syrian Archive exhibits the ingenuity that characterizes data activists; however, not all data activist initiatives employ the same data methods. Next, the commonalities in good enough data gathering practices are examined. First, Syrian Archive’s data are inspected through the lens of the criteria for ‘just good enough data’ (see Table 1).[[92]](#footnote-92) Second, based on the taxonomy of data gathering methods offered earlier,[[93]](#footnote-93) several data activist initiatives are compared with Syrian Archive looking at the variety of data mining approaches they employ (see Table 2). And third, how the interviewees themselves definite ‘good enough’ is explored to extract more insights.

## Comparison and discussion

### Meeting the criteria for ‘good enough’

First, Gabrys, Pritchard and Barratt refer to datasets generated by citizens, that is, ordinary people understood as non-experts.[[94]](#footnote-94) Meanwhile, although based on citizen data, the team at Syrian Archive includes ‘researchers, journalists, technologists and digital security experts who have been working in the field of human rights, verification, open source technologies and investigation methodologies for the past ten years’.[[95]](#footnote-95) According to the Syrian Archive’s protocols, experts determine the data rules, and ordinary people (citizen journalists and contributors) and other experts (conventional journalists) provide the data from the ground. At face value, this constitutes a difference. However, Gabrys, Pritchard and Barratt also acknowledge that expert mediation was required in citizen data sensing ‘given the disjuncture between the expertise needed to analyze the data and the researchers’ and residents’ skills in undertaking data analysis’.[[96]](#footnote-96) Namely, good enough data analysis involved citizens but requires a degree of expertise. However, the ‘good enough’ condition does not depend on distinctions between ordinary people and experts: data are either good enough or not. Gabrys, Pritchard and Barratt note ‘regulators, scientists and polluters’ have attempted ‘to discredit citizen data’ due to concerns about their know-how.[[97]](#footnote-97) Nevertheless, as they also say, ‘questions about validity do not pertain to citizen data alone’.[[98]](#footnote-98) Thus, the level of expertise of either data gatherers or data interpreters does not seem to be relevant as long as the data are robust.

Second, Syrian Archive provides new types of data and data stories, since no other actor, official or not, is producing them. Likewise, Gabrys, Pritchard and Barratt suggest that the citizens involved in the data sensing exercises are compelled to act by the absence of information and institutional support;[[99]](#footnote-99) that is, they are filling a gap. This is the case of Syrian Archive too; for instance, the official UN report on Aleppo shows only part of the story. This characteristic has to do with the ability of data activists of crafting alternative maps, stories and solutions.[[100]](#footnote-100) The purpose of filling gaps suggests that there is a particular amount of information needed to tell a human rights story accurately, and that activism can help secure this information. That is why good enough data implies to the right amount of data –not necessarily big— that activists needs to produce.

Third, as the Aleppo investigation suggests, the result of the analysis of the video footage was useful to generate new ‘forms of evidence’, paraphrasing Gabrys, Pritchard and Barratt,[[101]](#footnote-101) which were good enough to reveal a Russian or Syrian intervention in the bombing of the city. Thus, good enough data can be evidential data, since they can satisfy the need to provide sufficient support for ongoing legal investigations.

This paper proposes a new criterion. To be deemed good enough, data should also be suitable for the goals of the initiative in which the data are being employed, as Interviewee 3 suggests:

What is good enough data depends on the purpose of what you are doing (…) (For Syrian Archive) the dataset in question has to be the result of an investigation, it has to be cross-referenced, the product of a collaboration, it has to be in a standardized form if it comes from different platforms and media types, and it has to be preserved to be handed over for legal accountability and the end user, for example, a UN report.

It might be said that Syrian Archive’s long-term mission is to promote social change in Syria, but there are various means by which change might be promoted. Syrian Archive goes about fulfilling its mission in its own way: producing evidence of abuses that can withstand examination in court. The nature of what counts as ‘good enough’ data is *unavoidably* contextual. The data in the Syrian Archive are good enough to show that there are human rights violations in the military conflict. It appears we cannot ask whether a pertinent dataset is good enough without further formulating the specific use to which the data will be put; that is, without stipulating a) the specific role the data play in how change is to be fostered, and b) what change is to be advanced. The question of whether data are ‘good enough’ must be framed by the context in which the data are being used; it should respond to the question ‘what for’. And Syrian Archive’s data are good enough for the purposes of the organization, while they are new and are generating alternative data stories and evidence that can be the basis for court cases.

### How Syrian Archive fares in comparison with other cases

Data activists can resort to different data methods. Table s shows how Syrian Archive compares with the other cases mentioned before. The purpose of this table is to spell out the different origins, type of data activism and means to understand good enough data more broadly and provide a taxonomy that can be used as a tool to enhance the comparability of case studies.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Origins of data** | **Type of data activism** | **Means** |
| **Syrian Archive** | Videos on social media (posted by citizens and others) | 2nd (public data) and 3rd (crowdsourced data) | Videos uploaded by citizens, journalists and activists on sharing platforms and social media, news media; public satellite imagery; news media; testimonies from a network of sources |
| **Forensic Architecture** | Satellite imagery and data, ballistic analysis, news, citizen data | 2nd (public data)and 5th (own data) | Public satellite imagery and data (i.e. AIS signals); news media; sensors; a variety of means |
| **WeRobotics** | Drone-based imagery (gathered by citizens) | 5th (own data) | Drones; people as data/images analyzers |
| **InfoAmazonia** | Satellite imagery, crowdsourced citizen data, citizen sensing | 2nd (public data), 3rd (crowdsourced data) and 5th (own data) | Public satellite imagery; crowdsourcing platform; sensors (i.e. water quality) |
| **Ushahidi** | Crowdsourced citizen data; websites and social media | 2nd (public data)and 3rd (crowdsourced data) | Crowdsourcing platform and scraping methods (i.e. for verification); scraping |
| **Citizen sensing in Gabrys, Pritchard and Barratt (2016)** | Citizen sensing | 5th (own data) | Sensors |

*Table 1: Comparison of Data Initiatives by Their Origins[[102]](#footnote-102)*

Forensic Architecture employs advanced architectural and media research on behalf of international prosecutors, human rights organizations and campaigning groups to produce evidence of human rights abuses.[[103]](#footnote-103) WeRobotics launches ‘flying labs’ and swimming robots to capture data and images for social uses.[[104]](#footnote-104) Focused on promoting conservation via data transparency, InfoAmazonia visualizes journalistic stories and advocacy content on maps, sets up sensor networks to capture water data, creates alarm systems and crowdsources data from indigenous communities.[[105]](#footnote-105) Meanwhile, Ushahidi has created a crowdsourcing platform that gathers, verifies and visualizes citizen data to support humanitarian operations.[[106]](#footnote-106) One difference can be identified looking at their data sources is that, while the people who submit documents on social media may not know they are going to be used by Syrian Archive, the reporters of the Ushahidi platform and the citizens gathering data via sensors are creating data deliberately as part of the effort from its inception. The act of documenting a case of abuse and sharing the video may reveal an intention to denounce it and an acceptance that the information can be used by third parties, but it is not a given. Interviewees 3 and 4 say that some reporters send the footage directly to the organization; however, part of the Syrian Archive reporters’ involvement in the project is indirect.

In any case, these initiatives rely wholly or partially on citizens, and the interviews illustrate the importance of citizen data for their projects. Talking about Syrian Archive Interviewee 3 puts it like this:

(Syrian) footage gathered by ordinary citizens and journalists and posted online has started to appear as evidence in courtrooms. In August 2017, the International Criminal Court used a video posted on YouTube to get an international arrest warrant on an officer of the Libyan army accused of war crimes. Open information obtained from social media is increasingly recognized in court as potential evidence, which is really encouraging.

Interviewee 1 notes that the high value of citizen data in crisis mapping employed in digital humanitarianism:

Especially in times of disaster, the best people to give direction on what is needed on the ground are the people directly affected by the crisis, the otherwise assumed to be “passive recipients” of information.

Even the two non-activist interviewees integrate some forms of citizen data in their projects. Although the fragile states index does not incorporate them ‘because it is too difficult to verify at that high level’, social media or citizen data are part of ‘more ground-level, localized community assessments’, says Interviewee 5. The water grids designed to distribute water in cities and communities by Interviewee 2 employ user data to identify loss and malfunction within the grids.

However, there is a distinction to be made when the lens of volume and data roles are taken into account. In the case of Syrian Archive and Forensic Architecture, citizen data are incorporated on a case by case basis; WeRobotics, Ushahidi, InfoAmazonia and citizen sensing integrate citizens not only as data gatherers but also as analyzers and users.

Another trend that can be spotted is the alternative use practitioners make of data-harvesting technologies. For example, while drones - crewless aerial vehicles - were created originally for military purposes, they have been appropriated by WeRobotics and other data activists for social and humanitarian uses. Likewise, maps have been traditionally the monopoly of the state; these organizations and projects have seized them to generate alternative narratives and oppose top-down approaches.[[107]](#footnote-107) The same observations can be made about the data infrastructure being employed in activism, not its primary purpose. That is, these data practices not only generate new datasets and data stories; they also reverse top-down approaches to the data infrastructure.

### The interviewees: Verification and usefulness

What else can be said about good enough data? Next, the interviewees provide more clues as to what is good enough. For example, Interviewee 1 notes something that appears obvious: data have to be accurate to be valuable. To make sure this is the case, data practitioners resort to verification processes with a variety of techniques, which are, again, context related.

Verification mechanisms (…) vary from one deployment to another, depending on the context and intensity of the situation. Children mapping out the cost of chicken across the world will have a less rigorous verification process as compared to election monitors in Kenya. In most cases, verification will involve corroboration with trusted sources online, and on the ground, and mainstream media sources (…) I think the definition of good enough data is something that can only be determined from one deployment to another. Is it relevant to the topic of interest/goal of your deployment? Do the submissions give you enough information to act?

Interviewee 2 also highlights the fact that, while good enough data is a relative concept, all data have to be standardized and validated to be usable:

Good enough data is really a moving target. Each data submission (…) can lead to new, unsupported data that needs to be analyzed and added to our raw input data validation and normalization stage. We also need to have a data QA stage that extracts statistics that let us inspect incoming data and adjust accordingly. Our definition of good data could be: It passes our initial raw input data validation filters; it’s normalized to our system's internal requirements; it doesn’t produce abnormalities in data QA (quality assurance)statistics, and it fits into each of the system’s internal data structures.

Interviewees 1, 2 and 4 suggest that what is good enough today, can become *not so good* tomorrow, and that this is a work in progress. The interviewees imply too that achieving good enough data is a collective effort that requires meticulous team-work. Interviewee 4 adds the concepts of comparability and accessibility, as well as the requirement to integrate feedback to correct errors and avoid misinformation:

Good enough data are data that have been verified and offered in an accessible and standardized, easy manner, which can collaborate with other types of data. It is also important that they are accessible, so potential problems are fixed through feedback. We see people working regarding collecting data (on human rights) based on published data but without any criteria of verification. The risk there is to spread propaganda.

Syrian Archive devotes significant resources to verify not only the data but also the data sources. Likewise, Interviewee 5 points out the need to rely on credible sources to apply verification tests.

It is basically a reasonability test. In assessing 178 countries on such a broad breadth of indicators every year, it is impossible to directly verify every one of the 50 million or so data points that we analyze. To that degree, a certain amount of trust is necessary for the credibility of the sources included in the sample by the content aggregation process. The data may never be perfect; however, our process of triangulation with different types of data is a way of ‘balancing’ it and ensuring it is providing reasonable outputs.

The interviewees stress the demand for data to be enough in quantity and accurate, which requires verification processes that integrate corrections back into the system. Some of them use similar methods. For example, the triangulation in the vulnerable countries index resembles Syrian Archive’s process in that the latter includes three layers of data identification and confirmation as well. These interviews add new perspectives on the concept of good enough data: data should be standardized and comparable so they can endure verification processes, they should employ trustworthy sources and embed the capacity for absorbing corrections.

### Towards a definition of good enough data in activism

Building blocks that emerge from the comparisons, interviews and definitions above include: good enough data in activism can/should: 1) be robust enough in quantity and quality; 2) refer to forms of citizens’ involvement as data gatherers and other roles; 3) generate impact/action-oriented new data and data stories; 4) involve alternative uses of the data infrastructure and other technologies, and a variety of mining methods; 5) resort to credible data sources; 6) include standardization, comparability, accessibility, usability and accuracy, as well as verification, testing and feedback integration processes; 7) involve collaboration; 8) be relevant for the context and aims of the research questions; and 9) be preserved for further use, including as evidence in court cases.

The comparison between the reports on Aleppo issued by Syrian Archive and by OHCHR, and between activist and non-activist data gathering exercises show that data activism has potential to produce good enough data to generate dependable information, filling gaps, complementing and supporting other actors’ efforts and, paraphrasing Gabrys, Pritchard and Barratt,[[108]](#footnote-108) creating patterns of actionable evidence that can mobilize policy changes, community responses, follow-up monitoring, and improved accountability.

## Compliance with Ethics Requirements

The author declares that she has no conflict of interest. The author has permission to use the images included in the article. Informed consent was obtained from all interviewees for being included in the study.

# References

Al Jazeera News. ‘Syria’s Civil War Explained from the Beginning’, *Al Jazeera*, 14 April 2018. https://www.aljazeera.com/news/2016/05/syria-civil-war-explained-160505084119966.html.

Amnesty International. ‘Exposed: Child Labour behind Smart Phone and Electric Car Batteries’, 19 July 2016, https://www.amnesty.org/en/latest/news/2016/01/child-labour-behind-smart-phone-and-electric-car-batteries/.

———. ‘Syria: “War Of Annihilation”: Devastating Toll On Civilians, Raqqa – Syria’, 2018, https://www.amnesty.org/en/documents/mde24/8367/2018/en/.

Axinn, William G., Dirgha Ghimire and Nathalie E. Williams. ‘Collecting Survey Data during Armed Conflict’, *J Off Stat* 28.2 (2012): 153–171.

Balsiger, Philip, and Alexandre Lambelet. ‘Participant Observation’, in Donatella della Porta (ed.), M*ethodological Practices in Social Movement Research*, Oxford: Oxford University Press, 2014, pp. 144–72.

Banchik, Anna, et al. ‘Chemical Strikes on Al-Lataminah’. *Berkeley Human Rights Investigation Lab*, 25, 30 March 2017, https://syrianarchive.org/assets/hama/Syrian\_Archive\_Hama\_Report\_Final.pdf.

Berendt, Bettina, Marco Büchler, and Geoffrey Rockwell.‘Is It Research or Is It Spying? Thinking-Through Ethics in Big Data AI and Other Knowledge Science’, *Künstliche Intelligenz* 29.223 (2015), https://doi.org/10.1007/s13218-015-0355-2.

Bernard, H. Russell. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, Fourth Edition, Maryland: AltaMira Press, 2006.

Boellstorff, Tom. ‘Making Big Data, in Theory’, *First Monday* 18.10 (2013), http://firstmonday.org/article/view/4869/3750.

Borges, Jorge Luis. ‘Del Rigor En La Ciencia’, *Palabra Virtual*, 12 February 2007. http://www.palabravirtual.com/index.php?ir=ver\_voz1.php&wid=726&p=Jorge%20Luis%20Borges&t=Del%20rigor%20en%20la%20ciencia&o=Jorge%20Luis%20Borges

Braman, Sandra. 2009. *Change of State: Information, Policy, and Power*. Cambridge: MIT Press.

Cadwalladr, Carole, and Emma Graham-Harrison. ‘Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach’. *The Guardian*, 17 March 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Cohen, D., and B. Crabtree. ‘Qualitative Research Guidelines Project’. The Robert Wood Johnson Foundation: Semi-Structured Interviews. July 2006. http://www.qualres.org/HomeSemi-3629.html.

Cukier, Kenneth. ‘Big Data Is Better Data’. presented at the Berlin TEDSalon, Berlin, June 2014, http://www.ted.com/talks/kenneth\_cukier\_big\_data\_is\_better\_data?language=en

Dijck, Jose van. ‘Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology’. *Surveillance & Society* 12.2 (2014): 197–208.

Downing, John D. H. (ed.). *Encyclopedia of Social Movement Media*, Thousand Oaks: SAGE, 2011.

Ford, Heather. ‘Can Ushahidi Rely on Crowdsourced Verifications?’, *H Blog,* 3 September 2012, https://hblog.org/2012/03/09/can-ushahidi-rely-on-crowdsourced-verifications/.

Forensic Architecture. ‘About Saydnaya’, 2017, http://www.forensic-architecture.org/case/saydnaya/.

———. ‘Project’, 2028, http://www.forensic-architecture.org/project/.

Gabrys, Jennifer, and Helen Pritchard. 2015. ‘Next-Generation Environmental Sensing: Moving beyond Regulatory Benchmarks for Citizen-Gathered Data’, in AJ Berre, S Schade, and J Piera (eds), *Proceedings of the Workshop ‘Environmental Infrastructures and Platforms 2015*, Barcelona, pp. 57–65, http://ecsa.citizen-science.net/sites/default/ files/envip-2015-draft-binder.pdf.

Gabrys, Jennifer, Helen Pritchard, and Benjamin Barratt. ‘This Is the Other Aspect of “just” Good Enough Data –that Citizen Data Could Provide Ways of Realising Environmental and Social Justice’, *Big Data & Society* 1.14 (2016): 14.

Gangadharan, Seeta Peña. ‘Digital Inclusion and Data Profiling’. *First Monday* 17.5 (2012), http://firstmonday.org/article/view/3821/3199.

Gitelman, Lisa (ed.). *Raw Data Is an Oxymoron*, Cambridge MA/London: MIT Press, 2013.

Greenberg, Andy. ‘These Are the Emails Snowden Sent to First Introduce His Epic NSA Leaks’, *The Wired*, 13 October 2014, https://www.wired.com/2014/10/snowdens-first-emails-to-poitras/.

Gutierrez, Miren. *Data Activism and Social Change*, London: Palgrave Macmillan, 2018

———.‘Maputopias: Cartographies of Knowledge, Communication and Action in the Big Data Society – The Cases of Ushahidi and InfoAmazonia’. *GeoJournal* (2018): 1–20. https://doi.org/https://doi.org/10.1007/s10708-018-9853-8.

———‘The Public Sphere in Light of Data Activism’. *Krisis* (2018).

Hogan, Mel, and Sarah T. Roberts. 'Data Flows and Water Woes: An Interview With Media Scholar Mél Hogan'. 2015, http://bigdatasoc.blogspot.com.es/2015/08/data-flows-and-water-woes-interview.html.

Human Rights Council, ‘Report of the Independent International Commission of Inquiry on the Syrian Arab Republic’, UN General Assembly, 5 Februrary, 2013 http://www.ohchr.org/Documents/HRBodies/HRCouncil/CoISyria/A.HRC.22.59\_en.pdf.

InfoAmazonia. ‘THE ANNUAL CYCLES of the Indigenous Peoples of the Rio Tiquié 2005-2006'. 2016. https://ciclostiquie.socioambiental.org/en/index.html.

———. 2017. ‘About’. 2017. https://infoamazonia.org/en/about/.

Innerarity, Daniel. ‘Ricos Y Pobres En Datos’, *Globernance.org*, 22 February 2016, http://www.danielinnerarity.es/opinion-preblog/ricos-y-pobres-en-datos/.

International Telecommunication Union. ‘Measuring the Information Society Report’ 2014. Geneva: International Telecommunication Union (ITU). http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf.

Kitchin, Rob, and Tracey P. Lauriault. ‘Small Data, Data Infrastructures and Big Data’, *GeoJournal*, The Programmable City Working Paper 1, 80.4 (2014): 463–75.

Knight, Ben. ‘Syrian Archive Finds “Overwhelming” Russian Atrocities in Aleppo’. *DW*, 28 March 2017, http://www.dw.com/en/syrian-archive-finds-overwhelming-russian-atrocities-in-aleppo/a-38169808.

Korzybski, Alfred. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*, fifth edition, New York: Institute of General Semantics, 1994.

Lancaster, Lewis. ‘From Text to Image to Analysis: Visualization of Chinese Buddhist Canon’, *Digital Humanities,* 2010, http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/pdf/ab-670.pdf.

Lauriault, Tracey P.‘Data, Infrastructures and Geographical Imaginations’, 2012, Ottawa: Carleton University. https://curve.carleton.ca/system/files/etd/7eb756c8-3ceb-4929-8220-3b20cf3242cb/etd\_pdf/79f3425e913cc42aba9aa2b9094a9a53/lauriault-datainfrastructuresandgeographicalimaginations.pdf.

Leonelli, Sabina. ‘What Counts as Scientific Data? A Relational Framework’, *Philosophy of Science* 82.5 (2015): 810–21.

Lynch, Colum. ‘Soviet-Era Bomb Used in Syria Chemical Weapon Attack, Claims Rights Group’. *Foreign Policy*, 5 January 2017, http://foreignpolicy.com/2017/05/01/soviet-era-bomb-used-in-syria-chemical-weapon-attack-claims-rights-group/.

Lynd, Robert Staughton. *Knowledge for What: The Place of Social Science in American Culture*, Princeton: Princeton University Press, 1967.

Lyon, David. ‘Surveillance, Snowden, and Big Data: Capacities, Consequences’, *Big Data & Society* (July, 2014): 1-13.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston: Houghton Mifflin Harcourt, 2013.

Milan, Stefania and Miren Gutierrez. ‘Citizens´ Media Meets Big Data: The Emergence of Data Activism’. *Mediaciones* 14 (June, 2015). http://biblioteca.uniminuto.edu/ojs/index.php/med/article/view/1086/1027.

Miller, Harvey J. ‘The Data Avalanche Is Here: Shouldn’t We Be Digging?’ *Wiley Periodicals*, 22 August 2009.

Nakamitsu, Izumi. ‘The Situation in the Middle East’, Briefing S/PV.8344, New York: Security Council, 6 September 2018, https://www.securitycouncilreport.org/atf/cf/%7b65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7d/s\_pv\_8344.pdf.

Office of the UN High Commissioner for Human Rights. ‘Warring Parties Continued to Target Civilians for Abuses over Last Seven Months – UN Commission’, *United Nations,* 14 March, 2017, http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21370&LangID=E.

O’Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown Publishers, 2016.

Organisation for the Prohibition of Chemical Weapons. ‘Chemical Weapons Convention’, 2005, https://www.opcw.org/sites/default/files/documents/CWC/CWC\_en.pdf.

Palmer, Lindsay. ‘Ushahidi at the Google Interface: Critiquing the “geospatial Visualization of Testimony”’, *Continuum* 28.3 (2014): 342–56, DOI: https://doi.org/10.1080/10304312.2014.893989.

Poell, Thomas, Helen Kennedy, and Jose van Dijck. ‘Special Theme: Data & Agency’, *Big Data & Society*, 2 (July, 2015): 1-7. http://bigdatasoc.blogspot.com.es/2015/12/special-theme-data-agency.html.

Rogers, Simon. ‘John Snow’s Data Journalism: The Cholera Map That Changed the World’, *The Guardian*, 15 March 2012, http://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map.

Slater, Dirk. ‘What I’ve Learned about Mapping Violence’, *Fab Riders*, 3 April 2016, http://www.fabriders.net/mapping-violence/.

Solvang, Ole. ‘Key Finding on Use of Chemical Weapons in Syria’. Human Rights Watch, 2017, https://www.hrw.org/news/2017/10/05/key-finding-use-chemical-weapons-syria.

Stephens-Davidowitz, Seth. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, New York: Dey Street Books, 2017.

Syrian Archive. ‘About’, 2017, https://syrianarchive.org/en/about.

———. ‘Research Methodology’. 6 January 2017, https://syrianarchive.org/p/page/methodology\_digital\_evidence\_workflow/.

———. ‘Violations Database’, Database. 2016. https://syrianarchive.org/en/database?location=%D8%AD%D9%84%D8%A8%20:%20%D8%AD%D9%84%D8%A8&after=2016-07-01&before=2016-12-31.

———.‘Eyes on Aleppo: Visual Evidence Analysis of Human Rights Violations Committed in Aleppo - July - Dec 2016’, 2017, https://media.syrianarchive.org/blog/5th\_blog/Eyes%20on%20Aleppo.pdf.

Tufekci, Zeynep. ‘Engineering the Public: Internet, Surveillance and Computational Politics’, *First Monday* 19.7 (2014).

Ushahidi, ‘About’, 2017, https://www.ushahidi.com/about.

Vota, Wayan. ‘Dead Ushahidi: A Stark Reminder for Sustainability Planning in ICT4D’. *ICT Works,* 9 July 2012. http://www.ictworks.org/2012/07/09/dead-ushahidi-stark-reminder-sustainability-planning-ict4d/.

Weller, Marc. ‘Syria Air Strikes: Were They Legal?’ *BBC News*, 14 April 2018, https://www.bbc.com/news/world-middle-east-43766556.

WeRobotics, ‘About’, 2017, http://werobotics.org/.

Woodside, Arch. *Case Study Research : Theory, Methods and Practice*, United Kingdom: Emerald, 2010.

Yin, Robert K. *Case Study Research: Design and Methods,*Applied Social Research Methods Series. Thousand Oaks: SAGE,2002.

Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez and Vicente Chang. ‘Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints’, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* Copenhagen, Denmark, 7-11 September 2017, pp. 2979-2989.

1. Jennifer Gabrys, Helen Pritchard, and Benjamin Barratt. ‘This Is the Other Aspect of “just” Good Enough Data –that Citizen Data Could Provide Ways of Realising Environmental and Social Justice’, *Big Data & Society* 1.14 (2016): 14. [↑](#footnote-ref-1)
2. Stefania Milan and Miren Gutierrez, ‘Citizens’ Media Meets Big Data: The Emergence of Data Activism’, *Mediaciones* 14 (June, 2015): 120-133. [↑](#footnote-ref-2)
3. Miren Gutierrez, *Data Activism and Social Change*, London: Palgrave Macmillan, 2018. [↑](#footnote-ref-3)
4. See: Mel Hogan and Sarah T. Roberts, ‘Data Flows and Water Woes: An Interview With Media Scholar Mél Hogan’, *Big Data & Society,* 12 August 2015, http://bigdatasoc.blogspot.com.es/2015/08/data-flows-and-water-woes-interview.html; Lindsay Palmer, ‘Ushahidi at the Google Interface: Critiquing the “geospatial Visualization of Testimony”’, *Continuum* 28.3 (2014): 342–56; Wayan Vota, ‘Dead Ushahidi: A Stark Reminder for Sustainability Planning in ICT4D’, *ICT Works*, 9 July 2012, http://www.ictworks.org/2012/07/09/dead-ushahidi-stark-reminder-sustainability-planning-ict4d/ [↑](#footnote-ref-4)
5. See: Colum Lynch, ‘Soviet-Era Bomb Used in Syria Chemical Weapon Attack, Claims Rights Group’. *Foreign Policy*, 5 January 2017, http://foreignpolicy.com/2017/05/01/soviet-era-bomb-used-in-syria-chemical-weapon-attack-claims-rights-group/; Ole Solvang, ‘Key Finding on Use of Chemical Weapons in Syria’, *Human Rights Watch,* 5 October 2017, https://www.hrw.org/news/2017/10/05/key-finding-use-chemical-weapons-syria [↑](#footnote-ref-5)
6. Al Jazeera News, ‘Syria’s Civil War Explained from the Beginning’. *Al Jazeera*, 14 April 2018, https://www.aljazeera.com/news/2016/05/syria-civil-war-explained-160505084119966.html [↑](#footnote-ref-6)
7. Amnesty International, ‘Exposed: Child Labour behind Smart Phone and Electric Car Batteries’. 19 July 2016, https://www.amnesty.org/en/latest/news/2016/01/child-labour-behind-smart-phone-and-electric-car-batteries/ [↑](#footnote-ref-7)
8. Organisation for the Prohibition of Chemical Weapons, ‘Chemical Weapons Convention’, 2015 https://www.opcw.org/sites/default/files/documents/CWC/CWC\_en.pdf. [↑](#footnote-ref-8)
9. Marc Weller, ‘Syria Air Strikes: Were They Legal?’ *BBC News*, 14 April 2018, https://www.bbc.com/news/world-middle-east-43766556. [↑](#footnote-ref-9)
10. Izumi Nakamitsu, ‘The Situation in the Middle East’. Briefing S/PV.8344, New York: Security Council, 6 September 2018, https://www.securitycouncilreport.org/atf/cf/%7b65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7d/s\_pv\_8344.pdf. [↑](#footnote-ref-10)
11. Anna Banchik et al, ‘Chemical Strikes on Al-Lataminah’, *Berkeley Human Rights Investigation Lab*, 25, 30 March 2017, https://syrianarchive.org/assets/hama/Syrian\_Archive\_Hama\_Report\_Final.pdf. [↑](#footnote-ref-11)
12. Ibid. [↑](#footnote-ref-12)
13. Syrian Archive, About, https://syrianarchive.org/en/about. [↑](#footnote-ref-13)
14. Ibid. [↑](#footnote-ref-14)
15. Ibid. [↑](#footnote-ref-15)
16. Stefania Milan and Miren Gutierrez, ‘Citizens´ Media Meets Big Data: The Emergence of Data Activism’, *Mediaciones* 14 (June, 2015): 120-133. [↑](#footnote-ref-16)
17. Ibid. [↑](#footnote-ref-17)
18. Thomas Poell, Helen Kennedy, and Jose van Dijck, ‘Special Theme: Data & Agency’, *Big Data & Society*, 2 (July, 2015): 1-7. [↑](#footnote-ref-18)
19. Gutierrez, *Data Activism and Social Change*. [↑](#footnote-ref-19)
20. Ibid. [↑](#footnote-ref-20)
21. Ibid. [↑](#footnote-ref-21)
22. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’. [↑](#footnote-ref-22)
23. Ibid, 1. [↑](#footnote-ref-23)
24. Ibid, 11. [↑](#footnote-ref-24)
25. Arch Woodside, *Case Study Research: Theory, Methods and Practice*, United Kingdom: Emerald, 2010, p. 322. [↑](#footnote-ref-25)
26. Ibid. [↑](#footnote-ref-26)
27. Robert K Yin, *Case Study Research: Design and Methods,*Applied Social Research Methods Series. Thousand Oaks: SAGE,2002, p. 13. [↑](#footnote-ref-27)
28. Philip Balsiger, and Alexandre Lambelet, ‘Participant Observation’, in Donatella della Porta (ed.), M*ethodological Practices in Social Movement Research*, Oxford: Oxford University Press, 2014, pp. 144–72; Russell H. Bernard, *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, fourth edition, Maryland: AltaMira Press, 2006; D. Cohen and B. Crabtree, ‘Qualitative Research Guidelines Project’, The Robert Wood Johnson Foundation: Semi-Structured Interviews, July 2006, http://www.qualres.org/HomeSemi-3629.html. [↑](#footnote-ref-28)
29. Rob Kitchin and Tracey P. Lauriault, ‘Small Data, Data Infrastructures and Big Data’, *GeoJournal*, The Programmable City Working Paper 1, 80.4 (2014): 463. [↑](#footnote-ref-29)
30. Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston: Houghton Mifflin Harcourt, 2013. [↑](#footnote-ref-30)
31. Jose van Dijck, ‘Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology’, *Surveillance & Society* 12.2 (2014): 197–208. [↑](#footnote-ref-31)
32. International Telecommunication Union, ‘Measuring the Information Society Report’, *Geneva: International Telecommunication Union (ITU),* http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf, 2014. [↑](#footnote-ref-32)
33. Seth Stephens-Davidowitz*, Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, London: Bloomsbury Publishing, 2017. [↑](#footnote-ref-33)
34. Ibid. [↑](#footnote-ref-34)
35. Harvey J. Miller, ‘The Data Avalanche is Here: Shouldn’t we be Digging?’ Wiley Periodicals, Inc., 22 August 2009, 2. [↑](#footnote-ref-35)
36. Lewis Lancaster, ‘From Text to Image to Analysis: Visualization of Chinese Buddhist Canon’, *Digital Humanities*, 2010, pp. 1-3. [↑](#footnote-ref-36)
37. Kenneth Cukier, ‘Big Data Is Better Data’, presented at the Berlin TEDSalon, Berlin, June 2014, http://www.ted.com/talks/kenneth\_cukier\_big\_data\_is\_better\_data?language=en [↑](#footnote-ref-37)
38. Lisa Gitelman (ed.), *Raw Data Is an Oxymoron*, Cambridge MA/London: MIT Press, 2013. [↑](#footnote-ref-38)
39. Jorge Luis Borges, ‘Del Rigor En La Ciencia’, *Palabra Virtual*, 12 February 2007, http://www.palabravirtual.com/index.php?ir=ver\_voz1.php&wid=726&p=Jorge%20Luis%20Borges&t=Del%20rigor%20en%20la%20ciencia&o=Jorge%20Luis%20Borges [↑](#footnote-ref-39)
40. Ibid. [↑](#footnote-ref-40)
41. Alfred Korzybski, *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*, fifth edition, New York: Institute of General Semantics, 1994, p. 58. [↑](#footnote-ref-41)
42. Sabina Leonelli, ‘What Counts as Scientific Data? A Relational Framework’, *Philosophy of Science* 82.5 (2015): 810. [↑](#footnote-ref-42)
43. Tracey P. Lauriault, ‘Data, Infrastructures and Geographical Imaginations’, Ottawa: Carleton University, 2012, https://curve.carleton.ca/system/files/etd/7eb756c8-3ceb-4929-8220-3b20cf3242cb/etd\_pdf/79f3425e913cc42aba9aa2b9094a9a53/lauriault-datainfrastructuresandgeographicalimaginations.pdf. [↑](#footnote-ref-43)
44. Tom Boellstorff, ‘Making Big Data, in Theory’, *First Monday* 18.10 (2013), http://firstmonday.org/article/view/4869/3750. [↑](#footnote-ref-44)
45. Gitelman (ed.), *Raw Data Is an Oxymoron*. [↑](#footnote-ref-45)
46. Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Vicente Chang, ‘Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints’, *Association for Computational Linguistics* (2017): 2979–2989, DOI: https://doi.org/10.18653/v1/D17-1323. [↑](#footnote-ref-46)
47. Ibid. [↑](#footnote-ref-47)
48. See: David Lyon, ‘Surveillance, Snowden, and Big Data: Capacities, Consequences’, *Big Data & Society* (July, 2014): 1-13; Andy Greenberg, ‘These Are the Emails Snowden Sent to First Introduce His Epic NSA Leaks’, *The Wired*, 13 October 2014, https://www.wired.com/2014/10/snowdens-first-emails-to-poitras/; Bettina Berendt, Marco Büchler and Geoffrey Rockwell, ‘Is It Research or Is It Spying? Thinking-Through Ethics in Big Data AI and Other Knowledge Science’, *Künstliche Intelligenz* 29.223 (2015), DOI: https://doi.org/10.1007/s13218-015-0355-2. [↑](#footnote-ref-48)
49. Carole Cadwalladr and Emma Graham-Harrison, ‘Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach’, *The Guardian*, 17 March 2018, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. [↑](#footnote-ref-49)
50. Sandra Braman, *Change of State: Information, Policy, and Power*, Cambridge MA: MIT Press, 2009. [↑](#footnote-ref-50)
51. Zeynep Tufekci, ‘Engineering the Public: Internet, Surveillance and Computational Politics.’ *First Monday* 19.7 (2014), https://firstmonday.org/ojs/index.php/fm/article/view/4901/4097. [↑](#footnote-ref-51)
52. Seeta Peña Gangadharan, ‘Digital Inclusion and Data Profiling’, *First Monday,* 17.5 (2012) http://firstmonday.org/article/view/3821/3199. [↑](#footnote-ref-52)
53. Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown Publishers, 2016. [↑](#footnote-ref-53)
54. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’. [↑](#footnote-ref-54)
55. Cukier, ‘Big Data Is Better Data’. [↑](#footnote-ref-55)
56. Robert Staughton Lynd, *Knowledge for What: The Place of Social Science in American Culture*, Princeton: Princeton University Press, 1967. [↑](#footnote-ref-56)
57. Simon Rogers, ‘John Snow’s Data Journalism: The Cholera Map That Changed the World’, *The Guardian*, 15 March 2012, http://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map. [↑](#footnote-ref-57)
58. Daniel Innerarity, ‘Ricos Y Pobres En Datos’, *Globernance.org*, 22 February 2016, http://www.danielinnerarity.es/opinion-preblog/ricos-y-pobres-en-datos/. [↑](#footnote-ref-58)
59. Milan and Gutierrez, ‘Citizens’ Media Meets Big Data’. [↑](#footnote-ref-59)
60. See: Gutierrez, *Data Activism and Social Change*. [↑](#footnote-ref-60)
61. Forensic Architecture, ‘About Saydnaya’, http://www.forensic-architecture.org/case/saydnaya/. [↑](#footnote-ref-61)
62. WeRobotics, ‘About’, http://werobotics.org/. [↑](#footnote-ref-62)
63. InfoAmazonia, ‘The Annual Cycles of the Indigenous Peoples of the Rio Tiquié, 2005-2006’, 2016. https://ciclostiquie.socioambiental.org/en/index.html. [↑](#footnote-ref-63)
64. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’. [↑](#footnote-ref-64)
65. Ibid, p. 8. [↑](#footnote-ref-65)
66. John D. H. Downing (ed.), *Encyclopedia of Social Movement Media*. Thousand Oaks: SAGE, 2011, p. 18. [↑](#footnote-ref-66)
67. Gutierrez, *Data Activism and Social Change*; Miren Gutierrez, ‘Maputopias: Cartographies of Knowledge, Communication and Action in the Big Data Society – The Cases of Ushahidi and InfoAmazonia’. *GeoJournal*, (2018): 1–20, DOI: https://doi.org/https://doi.org/10.1007/s10708-018-9853-8; Miren Gutierrez, ‘The Public Sphere in Light of Data Activism’, *Krisis* (2018). [↑](#footnote-ref-67)
68. Dirk Slater, ‘What I’ve Learned about Mapping Violence’, *Fab Riders*, 3 April 2016, http://www.fabriders.net/mapping-violence/. [↑](#footnote-ref-68)
69. Wayan Voya, ‘Dead Ushahidi: A Stark Reminder for Sustainability Planning in ICT4D’, *ICT Works,* 9 July 2012. http://www.ictworks.org/2012/07/09/dead-ushahidi-stark-reminder-sustainability-planning-ict4d/. [↑](#footnote-ref-69)
70. Heather Ford, ‘Can Ushahidi Rely on Crowdsourced Verifications?’ *H Blog,* 3 September 2012, https://hblog.org/2012/03/09/can-ushahidi-rely-on-crowdsourced-verifications/. [↑](#footnote-ref-70)
71. Amnesty International, ‘Exposed: Child Labour behind Smart Phone and Electric Car Batteries’. [↑](#footnote-ref-71)
72. See: Gutierrez, *Data Activism and Social Change*; Hogan and Roberts, ‘Data Flows and Water Woes’; Lindsay Palmer, ‘Ushahidi at the Google Interface: Critiquing the “geospatial Visualization of Testimony”’, *Continuum* 28.3 (2014): 342–56, DOI: https://doi.org/10.1080/10304312.2014.893989. [↑](#footnote-ref-72)
73. Gutierrez, *Data Activism and Social Change*. [↑](#footnote-ref-73)
74. Syrian Archive, ‘Research Methodology’, 6 January 2017, https://syrianarchive.org/p/page/methodology\_digital\_evidence\_workflow/; [↑](#footnote-ref-74)
75. Ibid. [↑](#footnote-ref-75)
76. Office of the UN High Commissioner for Human Rights, ‘Warring Parties Continued to Target Civilians for Abuses over Last Seven Months – UN Commission’, *United Nations,* 14 March 2017, http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21370&LangID=E. [↑](#footnote-ref-76)
77. Ibid. [↑](#footnote-ref-77)
78. Forensic Architecture ‘Project’, http://www.forensic-architecture.org/project/. [↑](#footnote-ref-78)
79. William G. Axinn, Dirgha Ghimire and Nathalie E. Williams, ‘Collecting Survey Data during Armed Conflict’, *J Off Stat* 28.2 (2012): 153. [↑](#footnote-ref-79)
80. Ben Knight, ‘Syrian Archive Finds “Overwhelming” Russian Atrocities in Aleppo’. *DW*, 28 March 2017, http://www.dw.com/en/syrian-archive-finds-overwhelming-russian-atrocities-in-aleppo/a-38169808. [↑](#footnote-ref-80)
81. ‘Research Methodology’. 6 January 2017, p. 3, https://syrianarchive.org/p/page/methodology\_digital\_evidence\_workflow/. [↑](#footnote-ref-81)
82. Ibid. [↑](#footnote-ref-82)
83. Ibid. [↑](#footnote-ref-83)
84. Ibid. [↑](#footnote-ref-84)
85. Ibid. [↑](#footnote-ref-85)
86. Ibid. [↑](#footnote-ref-86)
87. Human Rights Council, ‘Report of the Independent International Commission of Inquiry on the Syrian Arab Republic’, UN General Assembly, 5 Februrary 2013, http://www.ohchr.org/Documents/HRBodies/HRCouncil/CoISyria/A.HRC.22.59\_en.pdf. [↑](#footnote-ref-87)
88. Source: Syrian Archive, ‘About’, https://syrianarchive.org/en/about. [↑](#footnote-ref-88)
89. Syrian Archive, ‘Research Methodology’, 19-20. https://syrianarchive.org/p/page/methodology\_digital\_evidence\_workflow/; ‘Violations Database’, https://syrianarchive.org/en/database?location=%D8%AD%D9%84%D8%A8%20:%20%D8%AD%D9%84%D8%A8&after=2016-07-01&before=2016-12-31. [↑](#footnote-ref-89)
90. Ibid. [↑](#footnote-ref-90)
91. Source:Syrian Archive, ‘Research Methodology’, https://syrianarchive.org/p/page/methodology\_digital\_evidence\_workflow/ [↑](#footnote-ref-91)
92. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’. [↑](#footnote-ref-92)
93. Gutierrez, *Data Activism and Social Change*. [↑](#footnote-ref-93)
94. See Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’. [↑](#footnote-ref-94)
95. Syrian Archive, ‘About’, https://syrianarchive.org/en/about. [↑](#footnote-ref-95)
96. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’, p. 6. [↑](#footnote-ref-96)
97. Ibid, p. 2 [↑](#footnote-ref-97)
98. Ibid. [↑](#footnote-ref-98)
99. Ibid. [↑](#footnote-ref-99)
100. Gutierrez, ‘Maputopias: Cartographies of Knowledge, Communication and Action in the Big Data Society’. [↑](#footnote-ref-100)
101. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’, p. 6. [↑](#footnote-ref-101)
102. Elaboration by the author based on Gutierrez*, Data Activism and Social Change*. [↑](#footnote-ref-102)
103. Forensic Architecture ‘Project’, 2028, http://www.forensic-architecture.org/project/. [↑](#footnote-ref-103)
104. WeRobotics, ‘About’, http://werobotics.org/ [↑](#footnote-ref-104)
105. InfoAmazonia, ‘About’, https://infoamazonia.org/en/about/. [↑](#footnote-ref-105)
106. Ushahidi, ‘About’, https://www.ushahidi.com/about. [↑](#footnote-ref-106)
107. Gutierrez,‘Maputopias: Cartographies of Knowledge, Communication and Action in the Big Data Society’. [↑](#footnote-ref-107)
108. Gabrys, Pritchard, and Barratt, ‘This Is the Other Aspect of “just” Good Enough Data’. [↑](#footnote-ref-108)