

## Challenges in Codifying Events Within Large and Diverse Data Sets of Human Rights Documentation: Memory, Intent, and Bias

JEFF DEUTCH<sup>1</sup>  
Syrian Archive, Germany

This article discusses challenges in codifying events within large and diverse data sets of human rights documentation, focusing on issues related to memory, intent, and bias. Clustering records by events allows for trends and patterns to be analyzed quickly and reliably, increasing the potential use of such content for research, advocacy, and accountability. Globally, archiving and preservation of user-generated digital materials documenting human rights abuses and war crimes are increasingly recognized as critical for advocacy, justice, and accountability. For the conflict in Syria, which began in 2011, there are more hours of user-generated content documenting rights violations uploaded to digital platforms than there have been hours in the conflict itself. Whereas some content has been clustered around specific events, such as larger open-source investigations by civil society and documentation efforts, the vast majority of content currently exists as individual unstructured records rather than jointly as clustered events within a relational database. The sheer amount of content and the near constant removals of materials from public channels mean that human rights monitors are in a race against time to preserve content, identify violations, and implicate potential perpetrators. Overcoming challenges related to memory and bias is crucial to this process.

*Keywords: preservation, open-source technology, open-source investigations, user-generated content, incident-based clustering, metadata, structured data, archiving, workflows, Syria*

Now entering its 10th year, the conflict in Syria has produced a massive amount of documentation concerning human rights violations and other crimes committed against the Syrian people, much of it

---

Jeff Deutch: [jeff@syrianarchive.org](mailto:jeff@syrianarchive.org)

Date submitted: 2017–12–04

<sup>1</sup> The documentation preserved, verified, and analyzed by the Syrian Archive has been collected by groups and individuals living or working in Syria. This would not be possible without their work documenting and reporting war crimes and human rights violations. I am thankful to those who have assisted in developing and refining this article, in particular members of the Syrian Archive team including Libby McAvoy, Hadi al Khatib, and Abdul Rahman al Jaloud. Additional thanks to the *International Journal of Communication* and the anonymous reviewers for their insightful comments. Finally, I would like to thank Dima Saber for organizing the Image Activism After the Arab Uprisings symposium.

Copyright © 2020 (Jeff Deutch). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

uploaded to the Internet. Syrian human rights defenders have a preference for using social media platforms for publishing and publicizing documentation or at least use the medium effectively and often (Sterling, 2012). Large social media platforms such as YouTube, Facebook, and Twitter have become “accidental archives” never intended to host such content. Yet, this open-source content often has the potential to offer counternarratives to misinformation and propaganda published by actors on all sides of the conflict. When citizen journalists upload verifiable video footage of an event, that footage might disprove claims made by those in traditionally powerful positions, such as states. But unfortunately, for various reasons, much of this important documentation has been erased or underutilized in advocacy and accountability work (al Jaloud, al Khatib, Deutch, Kayalli, & York, 2019).

To address this challenge, the Syrian Archive was founded in 2014 and has worked to preserve more than 3 million records of digital content from more than 3,000 sources, using this documentation to publish a number of investigations and long-form reports. The group aims to use this documentation to create a counternarrative to the misinformation created by all actors involved in the conflict in Syria, and to respond to human rights violations by introducing better tools, replicable methodologies, and a publicly available verified database for use in investigative journalism and reporting, evidence-based advocacy campaigns, and criminal case building to demand accountability against perpetrators of those violations.

Losing this documentation may directly affect justice and accountability efforts by Syrian regional and international civil society organizations. In some cases, this documentation might offer the only evidence that a war crime has happened. In an interview with *The Atlantic*, Syrian Archive founder Hadi al Khatib spoke to the ways in which losing this documentation will make it harder for groups to do their work. He stated,

[Cracking down on so-called extremist content deemed off limits in the West] could have ripple effects that make life even harder for those residing in repressive societies, or worse, in war zones. Any further crackdown on what people can share online . . . would definitely be a gift for all authoritarian regimes. . . . On the ground in Syria . . . Assad is doing everything he can to make sure the physical evidence [of potential human rights violations] is destroyed, and the digital evidence, too. The combination of all this—the filters, the machine-learning algorithms, and new laws—will make it harder for us to document what’s happening in closed societies. (Warner, 2019, para. 19)

Although preserving this content is important for justice and accountability efforts, the removal or erasure of this content also risks destroying the collective Syrian digital memory formed since 2011, which may result in ignoring violations committed by all parties in the conflict and prevent future generations from knowing what happened in their country.

Existing research by Syrian Archive researchers describes the data models used for storing and processing large data sets of human rights documentation (Deutch & Para, 2020). This includes processes of transforming diverse types of content from “original” online formats to “archival” formats accessible to researchers.

Further research by Syrian Archive researchers describes the operational models used to verify individual records (Deutch & Habal, 2018) for use in investigative pieces. In the case of videos uploaded to YouTube, for example, this involves knowing the source of a video, where a video was filmed, and when a video was filmed. Additional information regarding a video, such as landmarks present, the time of day, munitions observed, and others, aids in this process.

In this article, I discuss challenges related to archiving large and diverse data sets of human rights documentation, using the case study of the Syrian Archive's clustering of individual records into a relational database of "events." In the process, I focus on issues related to memory, intent, and bias. With millions of records preserved, contextualizing content into events would add an additional layer and increase the collection's potential to support and further justice and accountability efforts of all kinds. Clustering records by events allows for trends and patterns to be analyzed quickly and reliably, increasing the potential use of such content for research, advocacy, and accountability. Furthermore, clustering content into events could aid in identifying potential perpetrators as well as intent, both useful for advocacy and accountability.

I first discuss the use of imagery and "emotive affectivity" by advocacy organizations to pursue social justice aims, as well as data loss relating to records documenting human rights violations committed in the Syrian conflict. I then discuss the importance of archiving these records and the potential use of these records in legal proceedings. Finally, I discuss challenges in contextualizing these records, using the example of the Syrian Archive's clustering records into event-specific databases, as well as issues relating to bias of which types of violations are documented.

### **Memory and Machines**

There often exist tensions between authenticity and reliability of machine and human memory. Speaking to these tensions, Garcia (2016) writes,

Both human memory and machine memory are subject to the limitations, vulnerabilities, and maintenance requirements of its material medium, whether it is soft flesh or a hard drive. Everything fails, sooner or later. And, as foreseen in even the earliest of cyberpunk literature, machine memory is not inherently more truthful . . . it may be more precise, dense, and enduring in its recall of stored information, but the quality of that data is uncertain from the outset and subject to tampering and corruption. (p. 90)

In 1982, *TIME Magazine's* "Man of the Year"<sup>2</sup> was a machine: the computer. The title, which is given annually to "the greatest influence for good or evil," was meant to mark the start of the information age ("The Computer Moves in," 1983). That same year, Philip K. Dick's (1968) most famous novel *Do Androids Dream of Electric Sheep?* was used as the basis for Ridley Scott's science fiction noir-thriller *Blade Runner* (Deeley & Scott, 1982). The film, set in 2019 in a dystopian futuristic Los Angeles, follows the story of blade runner Rick Decker, a bounty hunter with the job of "retiring" six Nexus-6 androids that escaped from an off-world colony on Mars and returned to Earth.

---

<sup>2</sup> A title that goes to "the greatest influence for good or evil" changed in 1999 to "Person of the Year."

Ideas of memory are central in *Blade Runner*, with androids in the story having false memories implanted and not developing the level of complex emotions required to feel empathy. Characters often have a hard time remembering things clearly or, when they do, realize that the memories they thought they had were not their own. In the film version, photographs are used as artifacts that embody these false memories. Android characters refer to these photographs, comfortable in the knowledge that if there has been a picture taken, something actually happened and the memories are real:

"Look," [Rachel, one of the androids said to Deckard] "here's me with my mother." But Deckard knows better; he has his own tests for androids or "humanoid robots." . . . "Not your memories," Deckard had said to her, "but some else's," a "synthetic memory system" as fraudulent as the faked photo. Crushed, Rachel leaves him musing at his piano, flipping through another set of faked "old" snapshots he had commandeered from another android. He has also spread his own family snapshots on the piano top, some faded, browned, curling with age and use. These photos are presumably the real thing, true memories of a past that actually happened. (Trachtenberg, 2008, p. 112)

The viewer of course is left not knowing whether Deckard's own photographs, and by extension his memories, are any more real than those of the androids he seeks to retire. The withering of collective memory is also a theme in both the novel and film, with characters knowing that there was a nuclear war that rendered most of the Earth uninhabitable, sending most able-bodied humans to off-world colonies in the aftermath. But little is known of what led to the war, or even how it ended. That the film was released the same year as the Chernobyl disaster makes that fear all the more real.

It is fitting that photographs are used as the physical embodiment of memories, as imagery has been shown to be very effective at evoking empathy in people who often have involuntary emotional responses when presented with imagery (Bernat, Benning, & Tellegen, 2006).

Moving from the world of science fiction to the real-world examples of science, those documenting conflicts have, since the invention of photography, relied on emerging technology as a means of "witnessing," to evoke a sense of realness, understanding, and objectivity. Webb Williams (2015) speaks to this when she discusses the role of witnessing and modern journalism:

We place value on reports from individuals on the ground: "The intrinsic value of 'being there,' on the ground, has been prized since the earliest days of crisis journalism" (Allan, 2013, pp. 9–10). Witnessing blends together subjectivity and objectivity. On the one hand, the witness (whether a professional journalist or not) has first-hand experiences to relate. On the other, they are subject to "the tensions besetting human understanding, interpretation, and memory" (Allan, 2013, p. 10). Media consumers, presumably, value this blend of objectivity and subjectivity in reporting. Eyewitness reports have an "emotive affectivity" that lends them the power to move others. (Webb Williams, 2015, p. 11)

These concepts hold true also for civil society organizations, which often use emotive affectivity in advocacy efforts. In a legal environment, courts overwhelmingly rely on witness statements and testimony

more than other types of documentation or forms of evidence. Although this has the potential to empower those affected, witness statements are known to be unreliable (National Research Council, 2014).

Users of social media platforms often face difficulties in knowing who and what to trust, or why and how to trust it. The promises of social media platforms as a force for good and meaningful connections are increasingly contrasted with an image of dystopian data brokers: collecting, processing, and selling personal information for monetary and/or political gain. Social media profiles are hacked by state-sponsored cyber armies to surveil users or remove content. Bots and fake profiles are used widely to steal information or to promote misinformation, propaganda, or conspiracy theories. So-called "fake news" is used to influence elections (Cadwalladr & Graham-Harrison, 2018), promote genocide (Mozur, 2018), recruit terrorists, and enable harassment (Pringle, 2017). An industry has risen whereby state and nonstate actors hire commercial third parties to carry out their technological dirty work, as evidenced by the hiring of the Israeli spyware company NSO Group, the Italian group Hacking Team, and others by various repressive states to target human rights defenders and journalists.

Inaccuracy of online data can have real implications in the offline world. For example, people often have similar names and records can accidentally be combined. When law enforcement relies on inaccurate data sets, a case of mistaken identity could lead to wrongful arrest or use of significant force against the wrong person. Indeed, former director of the NSA and the CIA General Michael Hayden has publicly stated, "We kill people based on metadata" (Cole, 2014, para. 2).

Erased or unavailable data present another challenge. There is growing concern and discourse concerning the unbridled power that social media platforms have in making editorial decisions of what is published for such large, global, nonheterogeneous audiences. Widespread takedowns of content on platforms that document human rights violations introduce challenges related to scale and false positives.

Starting in 2017, under rising pressure largely from Western governments to remove so-called "extremist content," platforms including YouTube and others started using machine learning technology to flag violating content on their platform for review by their teams (Google, 2020). For example, in 2019 YouTube removed 31.9 million videos, roughly 87,000 per day. Of those flagged for potential violation of terms of service, 87% were removed through automated flagging, and more than one third were removed before any views. Facebook removed roughly 25 million pieces of content deemed "terrorist propaganda" in 2019. In the first quarter of 2020, 99.3% of content was removed before users had reported it (Facebook, 2020). Twitter removed 115,861 accounts for terrorist content in the first half of 2019 (Twitter, 2020).

This process was responsible for terminating thousands of YouTube channels of groups that were publishing hundreds of thousands of videos documenting human rights violations. In Syria, these included groups such as the Syrian Observatory for Human Rights, the Violation Documentation Center, Sham News Agency, Aleppo Media Center, and many others. Content moderation is not limited to Syria, but also occurs elsewhere in conflict and nonconflict areas alike (al Jaloud et al., 2019). In Syria, many of the terminated social media accounts received strikes from YouTube, stating that videos violated its community guidelines by publishing violent/graphic content or by publishing content that incited violence or encouraged dangerous activities.

For this reason, the Syrian Archive has, since 2017, analyzed records within its collection, identified affected channels and accounts, and worked with other civil society organizations (e.g., Witness) as well as platforms (e.g., YouTube) to reinstate hundreds of thousands of records. Roughly 343,339 YouTube videos and 299 YouTube channels that the Syrian Archive holds in its collection were unavailable publicly on YouTube as of May 2020, amounting to 19.89% of videos and 8.57% of channels. This rate roughly doubled in just one year, with 13.04% of content in its collection publicly unavailable elsewhere in April 2019.

Content is also removed on Twitter: 127,719 tweets and 96 Twitter feeds were publicly unavailable on Twitter as of the date of publication, amounting to 12.51% of tweets and 19.05% of Twitter feeds (Syrian Archive, 2020a). This rate roughly tripled in just one year, with 6.78% of content in its collection publicly unavailable elsewhere in April 2019. It is entirely likely that “iconic images” of the conflicts in Syria and elsewhere that have been published to social media platforms have already been deleted from those platforms. This is particularly the case if the source is not alive, is arrested, or does not have access to e-mail, all incredibly common issues in a conflict area such as Syria.

Archiving these records is important for legal contexts. In legal environments, there is an emerging body of case law in which user-generated content from social media platforms features prominently, which is instructive. For example, in 2016 in Sweden, a case was concluded against a former Syrian rebel who had previously taken part in the killing of seven captured Syrian soldiers (Anderson, 2017). The court relied on content published on Facebook and Twitter to identify the time and place where soldiers were captured, as well as that only 41 hours had passed between their capture and execution. Prosecutors contacted Facebook to verify the content’s metadata.

Furthermore, in 2017 and again in 2018, the International Criminal Court issued an arrest warrant for a Libyan national by the name of Mahmoud Mustafa Busayf Al-Werfalli (*The Prosecutor v. Mahmoud Mustafa Busayf Al-Werfalli*, 2017). Werfalli was accused of being directly responsible for killing 33 people, based largely on video material and transcripts of video material posted to social media.

There often exists, however, a tension between more traditional legal institutions (e.g., truth commissions and courts) and civil society efforts such as journalists and nongovernmental organizations. Kohli (2018) writes that legal institutions are not always willing to share information with journalists, fearing that doing so might impact the conviction of a perpetrator in a court. But in conflict areas, she notes that “the government or administration might not be fit for the task of investigating human rights violations on its own, or worse—be itself directly involved in the abuses” (p. 68) This is particularly true in conflict areas such as Syria, where foreign journalists, nongovernmental organizations, and international monitoring agencies face difficulties accessing the country to document human rights violations (Higgins, 2013).

Open-source documentation, particularly that comprising user-generated content, will likely not replace more traditional forms of evidence gathering such as the collection of witness testimony and physical documentation. Rather, the aim of open-source investigations is to supplement more traditional forms of evidence gathering, particularly when used in a legal context.

Owens (2019) refers to this as a tension between information suppliers and demanders. Suppliers include those who collect, preserve, and distribute information (e.g., the Syrian Archive and other documentation efforts), and demanders refer to prosecutors and others who must navigate the complex web of suppliers. Owens proposes a regulatory framework that seeks to ensure that suppliers provide credible and reliable information that has been gathered in an ethical manner and that demanders will have more efficient access to that information. She notes that the current self-determination of applicable standards by suppliers means that "information gathered may be of no use to, or even hinder, future accountability efforts" (pp. 370–371). Despite these strides, courts and traditional documentation groups can lag behind in using the digital tools and methodologies required to harness this potential.

### **Contextualizing Content**

Preservation of materials documenting human rights violations is only one of the challenges posed to documentation groups. Finding, accessing, and making sense of relevant content and adding context to transform it into useful information present another (Kelley, 2002; von Krogh & Spaeth 2007). With hundreds of videos uploaded per day to potentially thousands of different channels on multiple platforms, often in nonstandardized formats lacking metadata and differing by platform (Embedded Metadata, 2015; Rieks, 2013), the context in which content is uploaded is inherently lost. This creates difficulties in linking multiple units of digital content of the same event, where a unit in this context refers to, for example, an individual video uploaded to YouTube or a posting uploaded to Facebook.

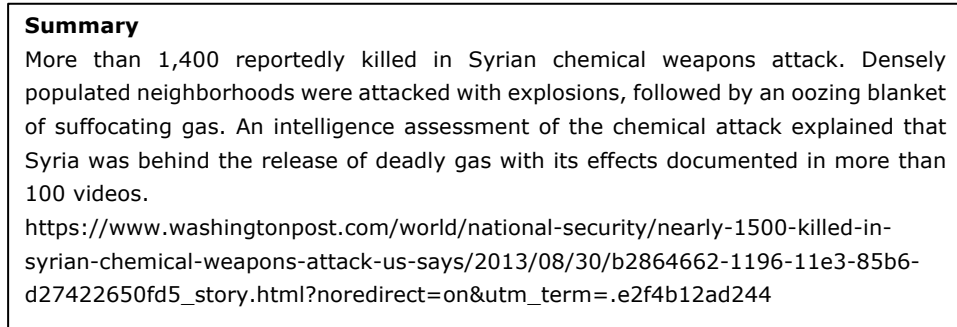
Looking at how the Syrian Archive approaches defining an "event" is useful for illustrating challenges in contextualizing these records. It is also useful for demonstrating the ways in which bias is present in annotation and clustering of digital content. A review of literature on event identification on social media shows that a news event is defined both in terms of geographic and temporal locality: something that happens in a specific place at a specific time that includes media coverage (e.g., Aggarwal & Subbian, 2012; McMinn et al., 2013; Panagiotou et al., 2016). Weng and Lee (2011) and Boettcher and Lee (2012) expand this definition to define events not by the media coverage a particular happening or occurrence receives, but by their shared topics in a given time period.

However, detecting unique online and offline events using user-generated content is difficult. The sheer quantity of content being created "in the wild," much of which is unrelated, unhelpful, or "noise," has led several researchers and academics to investigate the use of machine learning to automate this process (e.g., Li et al., 2012; Panagiotou et al., 2016; Papadopoulos et al., 2011; Petrovic et al., 2010).

Fortunately, thanks to the efforts of documentation groups such as the Syrian Archive, which provide verified data sets of individual records, researchers do not have to start from scratch and can limit the scope of a particular query to date, location, and keyword. Syrian Archive researchers have found that assigning a human-mediated discourse identifier in the form of an event-specific incident code is useful for tagging related records bound in temporal and geographic scope.

The Syrian Archive assigns descriptive titles to events, and summaries are written that include relevant mentions of the incident in the press or by other human rights monitoring groups with the goal of

providing an overview of the incident and of the quality of the available documentation. Elements in the summary include when the incident happened (date and time), the type of incident (e.g., explosion, aerial bombardment, chemical attack; targeting a hospital, bakery, school, etc.), location of the incident, those killed or injured, the type of munitions allegedly used in the attack, and whether there is any visual documentation showing the remnants of munitions (see Figure 1).



**Figure 1. Example of a Syrian Archive event summary.**

For each event-identification code generated, a detailed process of discovery is initiated to find additional documentation. This process necessitates internal and external discovery to identify records already within Syrian Archive infrastructure using date, location, and keyword searches, as well as using content discovery tools for social media content that is not in the archive.

Through adding context to data already existing within Syrian Archive infrastructure, this process transforms digital content to useful and searchable information with querying capabilities. Additional open-source software, such as VFRAME (<https://vframe.io>), a deep convolutional neural network object-detection tool developed by artist and technologist Adam Harvey, is used to automate identifying and triaging relevant content in pictures and videos trained on photorealistically rendered synthetic data. There is currently no open-source and publicly available library of munitions for machine learning training purposes; therefore, Syrian Archive staff have worked with VFRAME to develop one.

Yet, even within a specific temporal and geographic bound, further restrictions in temporal scope to set a minimum and maximum time interval and the presence or absence of additional features are necessary to accurately define an event as well as to introduce proxies for codifying features, such as intent.

Like codifying an event, intent is not something that can easily be observed or extracted by just looking at data. Although in some cases, notably the Al-Werfalli case mentioned earlier, an alleged perpetrator speaks directly to the camera and makes all intentions of committing a human rights violation known, most documentation of human rights violations is so-called "crime-based," meaning the documentation details the effects only after a crime has already been committed. Because perpetrator documentation in most cases is not available, it becomes difficult to indicate potential intent, making the difference between a deliberate crime and an unfortunate but legitimate accident unclear.



Although recognizing the limitations of purely technical automation to reliably annotate sensitive information, I propose that it is possible to codify potential intent through a variety of proxies. Through taking a big-data approach, groups such as the Syrian Archive include in their database *double-tap attacks* in which a protected object (e.g., schools or hospitals) is directly hit and receives multiple targeted strikes thereafter. The Syrian Archive defines a *direct hit* as when munitions or missiles hit a protected object, resulting in physical impact on the structure used for protected activities. On the other hand, the Syrian Archive considers an attack an *indirect hit* when munitions or missiles strike nearby buildings and the resulting explosion causes minor damage (i.e., not enough for critical structural harm) to the protected object or when the impact site for the attack is somewhere other than the protected object (e.g., an impact crater in the street in front of a hospital).

In the case of Syria, *double tap* refers to an attack on a particular target with a follow-up attack several minutes later to target first responders. For the Syrian Archive, *double tap* refers to multiple, distinct strikes at strategically relevant locations and at time intervals that indicate that the subsequent strikes are intended to cause harm to the responding humanitarian personnel. Through codifying whether an attack is a double tap or not, it becomes possible to identify the use of an attack tactic that may illustrate the perpetrators' intent to target first responders.

For this purpose, the Syrian Archive defines a double-tap strike as including geographic and temporal localities, as well as the type of munitions and delivery method. A double-tap event, therefore, must include more than one strike, a minimum time interval of seven minutes, and a maximum time interval of three hours. This demonstrates that the perpetrator using this tactic allowed sufficient time for humanitarian personnel to arrive on the scene before launching subsequent strikes. If there was little to no time between strikes, this may instead qualify as a different type of attack. The maximum time interval helps to ensure that the multiple strikes are linked in purpose. The event must also have subsequent strikes targeted either at the same site or to strategically linked site(s), each strike must be distinct (to exclude sustained fire and indiscriminate bombing), and small arms or light weapons are not used in the strikes.

Double taps are not limited to two strikes but could be many strikes so long as the targeting strikes are distinct, fit within the qualifying time intervals, and cannot be described as sustained fire. The same double-tap attack could also use multiple or varying delivery methods and munitions. For example, a qualifying use of the double-tap tactic occurs when a targeted site is first bombed from a helicopter and then struck by rocket artillery at least seven minutes later.

Finally, researchers at the Syrian Archive use the term *multiple targeted strikes* to define a tactic used to ensure successful targeting and maximum damage to a site in that there are multiple strikes on the same site in quick succession. The team defines a spatial margin of error (an approximately 50-meter radius) and certain timeframe (approximately three hours) as a proxy to demonstrate intent to hit and ensure significant damage to protected persons, objects, or sites that exist within that targeted strike site.

Although the double-tap and multiple targeted strikes tactics are distinct from a direct hit, several or all might apply to an event, allowing researchers to define events even within a particular temporal or

geographic bounding as distinct, adding an additional layer of refinement that might prove useful in accountability work. These are not the only categories that might prove useful for identifying intent; others, such as the proximity of attacks to frontlines, might indicate whether a protected person or place was intentionally targeted.

Definitions of events described earlier in this article (e.g., Li et al., 2012; Panagiotou et al., 2016; Papadopoulos et al., 2011; Petrovic et al., 2010) offer helpful starting points for event detection, but lack the contextual specificity required by documentation groups and those working toward accountability to use open-source content efficiently and effectively to supplement more traditional forms of evidence gathering, such as the collection of witness testimony and the collection of physical documentation.

Ultimately, these distinctions are hard to make and a combination of quantitative and qualitative research is necessary to annotate accurately. Therefore, a purely technical solution is not feasible and a mixed-methodology approach is necessary. In the end, rather than take a "one tool fits all" approach, in which workflows and methodologies are determined and limited by the use and design of technology tools, a more modular workflow in which different tools and methods are used at different stages of the process should be used for annotating large sets of event-specific content. It must be recognized, however, that even in the Syrian Archive's aim to remain impartial and unbiased, the objectiveness of clustered data and event criteria can be debated.

### **Bias and Neutrality of Archives and Archival Activist Research**

Various scholars have questioned the idea of an archive being neutral or objective. Schwartz and Cook (2002) stress that archives "are not passive storehouses of old stuff, but active sites where social power is negotiated, contested, confirmed" (p. 1). And Briones (2019) writes that archives are "power devices that create realities by storing and making pieces of evidence accessible" (p. 171). Purposefully or not, the archive introduces biases and subjectiveness. Similarly, in the worlds of data and statistics, Desrosières (2002) writes about how the history of statistics is the history of the state. When data are collected, he writes, it often takes the form of reifying dominant power structures.

In *The Archaeology of Knowledge: And the Discourse on Language*, Foucault (2010) defines the archive as not "the library of libraries"; nor is it "the sum of all the texts that a culture has kept upon its person as documents attesting to its own past" (p. 129), but rather "the systems of discursivity" that establishes the possibility of what can be said. Foucault uses the example of academic disciplines as systemic conceptual frameworks that define their own truth criteria. As Eliassen (2011) describes, Foucault sees an "archive" as being in conceptual symmetry with the concept of "discourse." Eliassen writes, "While discourse refers to the production of statements, archive indicates the selection of them" (para. 5). Manoff (2004) similarly writes that "archival work is about making fine discriminations to identify what is significant from a mass of data" (p. 19). Determining how to make discriminations of what is important or not from this mass of data is an issue central to the work efforts of the Syrian Archive.

McCracken (2017) argues that an archival activist approach differs from more traditional archives as described by Foucault (2010) and Eliassen (2010) in that archival activism is "strongly rooted in grassroots

activism, documenting social inequality and human rights” (para. 6). Briones (2019) explains that archival activism often serves two functions: (1) It offers access to stories, evidence, facts, and arguments that can be used to advance causes and social campaigns; and (2) it can be considered a counterculture practice itself, archiving material that has not been recognized by the dominant official structures.

If we apply this to the objectives, workflows, and methodologies used by the Syrian Archive, the group is not simply creating a “mirror” or a copy of YouTube or Facebook; nor is it making an external backup of all content being produced by Syrian people on those platforms. Rather, it is selective of which content is preserved based on a number of criteria. Furthermore, the group is selective about which content is verified and made publicly available on its own technical infrastructure. Choosing to be selective limits the amount of content that can be catalogued and published. Choosing to focus on documentation of human rights violations limits types of media that are catalogued and published. This inherently introduces bias into the work.

Of course, there are many forms of bias that inform the work of the Syrian Archive, just as in all documentation groups. None of the sources used for content are objective. Although efforts are made to review whether a source is reliable and not simply republishing existing content, each source has an agenda. For this reason, it is important to never rely on one source to reach a conclusion in an investigation. Although it is very possible that one source can create, even inadvertently, misinformation, it is highly unlikely that many groups reporting on the same incident have done so.<sup>3</sup>

For the Syrian Archive, as with any archival or data group, one always runs the risk of using categories or fields that inadvertently reify or reproduce dominant power structures; at best, this may be unnecessary, at worst this may be harmful. Although the data ontology used by the Syrian Archive includes hundreds of fields (Deutch & Para, 2020), many are quite subjective. To address this issue, the Syrian Archive takes an unstructured data approach that captures content as it was seen “in the wild” to allow both those within the team and others to recategorize, group, or cluster data.

To give one example, violation categories used in the Syrian Archive metadata schema include those recognized by the United Nations Office for High Commission of Human Rights (Syrian Archive, 2020b). This was a decision made early on to allow findings and documentation materials to be most useful to those in a position to implement change. But, of course, it limits the types of violations investigated and is bound by the UN’s definitions of what violations are.

In addition, certain types of human rights violations are more visible than others; as a result, they are more likely to have visual documentation. These include, for example, film footage documenting the use of illegal munitions, the destruction of civilian infrastructure, or the targeting of specifically protected persons and objects, such as hospitals (McDermott, Murray, & Koenig, 2019). A cursory look through available documentation on the Syrian Archive’s website has virtually none of sexual or gender-based

---

<sup>3</sup> Of course, as much as the Syrian Archive staff attempts to be impartial, everyone comes with his or her own set of personal biases and this is mostly likely reflected in the work of the Syrian Archive.

violence, very few of torture aside from what has been made public with leaks, and virtually nothing on displacements or forcible evictions.

This could be in part because of the networks the Syrian Archive is engaged in, the keywords the organization is searching for, the places staff is looking, and which forums these images are being shared in. Although it is possible that images of the effects of torture are being shared in private WhatsApp conversations or groups, in many cases the Syrian Archive does not have access to this.

What is important for a documentation group such as the Syrian Archive to remember is that just because there is bias at any stage of the process, it does not mean that content cannot be used. In addition, even information coming from biased sources (e.g., an alleged perpetrator) or content grouped using biased event criteria can be used to further social justice aims.

### **Concluding Remarks**

With so much content documenting human rights violations published and removed on a daily basis, documentation groups such as the Syrian Archive need to act fast in an organized and systematic way to securely preserve, verify, cluster, and publish information. There exist many challenges to harnessing the potential of open-source content for human rights research. Chief among these are issues of verification of content, an issue that has been written about extensively elsewhere (e.g., Daragahi, 2014; Browne, 2017; Cadwalladr, C. & Graham-Harrison, E., 2018, Gutierrez, 2019, Owens, 2019). The secure preservation of content and loss of data and memory are equally important and have been written extensively by archivists and nonarchivists alike (e.g., Asher-Shapiro, 2017; Baladi, 2016; Day, 2006; Graham, 2017).

Although discourses on archiving in the social sciences have existed for many years, their application to a human rights context in the digital realm is relatively new. Debates on data and objectivity of ample documentation must be negotiated with questions of not knowing how that documentation was made. Foucault's (2010) linking of archiving to power makes us question these further, with the introduction of biases and subjectivity. A multiplicity of truths, framed through a variety of disciplinary lenses, introduces more questions than answers and problems for the accurate reporting of unfolding events. A balance must be struck between recognizing these tensions and the need for human rights communities to highlight social injustices.

Further research is needed to expand the use of machine learning and object and scene identification to increase efficiency of internal discovery processes, as well as to incorporate additional types of data, such as documents, flight, and marine traffic data.

Although the fields of data activism, archival activism, and open-source investigations for human rights research have been expanding rapidly in recent years, only a small amount of such efforts have been documented in existing research, and few frame these issues in terms of the tensions of bias and memory in event definition. It was my objective in this article to address some of these, and I hope that this article will be helpful to other archival and investigative practices by civil society groups, journalists, and human rights researchers.

### References

- Aggarwal, C., & Subbian, K. (2012). Event detection in social streams. *Proceedings of the 2012 SIAM International Conference on Data Mining*, pp. 624-635. Anaheim, CA.
- al Jaloud, A. R., al Khatib, H., Deutch, J., Kayalli, D., & York, J. C. (2019). *Caught in the net: The impact of "extremist" speech regulations on human rights content*. Retrieved from <https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content>
- Anderson, C. (2017, December 11). Syrian rebel gets life sentence for mass killing caught on video. *The New York Times*. Retrieved from <https://www.nytimes.com/2017/02/16/world/europe/syrian-rebel-haisam-omar-sakhanh-sentenced.html>
- Asher-Schapiro, A. (2017). YouTube and Facebook are removing evidence of atrocities, Jeopardizing cases against war criminals. *The Intercept*. Retrieved from <https://theintercept.com/2017/11/02/war-crimes-youtube-facebook-syria-rohingya/>
- Baladi, L. (2016). Archiving a revolution in the digital age, Archiving as an act of resistance. *Ibraaz*. Retrieved from <http://www.ibraaz.org/essays/163>
- Bernat, E., Benning, C. J., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, 43(1), 93-103.
- Boettcher, A. & Lee, D. (2012). EventRadar: A real-time local event detection scheme using Twitter stream. In *2012 IEEE International Conference on Green Computing and Communications*, pp. 358-367. Washington D.C.
- Briones, M.A. (2019). A taxonomy of data visualization projects for alternative narratives. In L. Rampino & I. Mariani (Eds.), *Advancements in design research: 11 PhD theses on design as we do in Polimi, FrancoAngeli journals* (pp. 282-318). Milan, Italy: FrancoAngeli. Retrieved from [http://ojs.francoangeli.it/\\_omp/index.php/oa/catalog/book/376](http://ojs.francoangeli.it/_omp/index.php/oa/catalog/book/376)
- Browne, M. (2017). YouTube removes videos showing atrocities in Syria. *The New York Times*. Retrieved from <https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos.html>
- Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. Retrieved from <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Cole, D. (2014, May 10). *We kill people based on metadata*. Retrieved from <http://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/>

- The computer moves in. (1983, January 3). *TIME*. Retrieved from <http://content.time.com/time/subscriber/article/0,33009,953632-3,00.html>
- Day, M. (2006). The long-term preservation of Web content. In J. Masanès (Ed.), *Web archiving* (pp. 177–199). Berlin and New York: Springer.
- Daragahi, B. (2014). Misinformation maligns Syrian uprising. *Financial Times*. Retrieved from <http://blogs.ft.com/the-world/2014/04/misinformation-maligns-syrian-uprising/>
- Deeley, M. (Producer), & Scott, R. (Director). (1982). *Blade runner*. United States: Warner Bros.
- Desrosières, A. (2002). *The politics of large numbers: A history of statistical reasoning*. Cambridge, MA: Harvard University Press.
- Deutch, J., & Habal, H. (2018). The Syrian Archive: A methodological case study of open-source investigation of state crime using video evidence from social media platforms. *State Crime Journal*, 7(1), 46–76.
- Deutch, J., & Para, N. (2020). Targeted mass archiving of open source information: A case study. In S. Dubberly, D. Murray, & A. Koenig (Eds.), *Digital witness: Open source investigations for human rights advocacy and accountability* (pp. 165–184). Oxford, UK: Oxford University Press.
- Dick, P. K. (1968). *Do Androids dream of electric sheep?* New York, NY: Random House.
- Embedded Metadata. (2015). Embedded media manifesto. Retrieved from <http://www.embeddedmetadata.org/social-media-test-results.php>
- Eliassen, K. O. (2011). Archives and heterotopias. *Kunstjournalen B-post*, 10/11. Retrieved from <https://kunstjournalen.no/en/11/eliasen.html>
- Facebook. (2020). *Community standards enforcement report*. Retrieved from <https://transparency.facebook.com/community-standards-enforcement#dangerous-organizations>
- Foucault, M. (2010). *The archaeology of knowledge: And the discourse on language*. New York, NY: Vintage Books.
- Garcia, L. M. (2016). Withered memories and the ethnography of hidden things. *Continent*, 5(1), 83–91.
- Google. (2020). *YouTube community guidelines enforcement*. Retrieved from <https://transparencyreport.google.com/youtube-policy/removals>
- Graham, P. (2017). Guest editorial: Reflections on the ethics of Web archiving, *Journal of Archival Organization*, 14(3–4), 103–110.

- Gutierrez, M. (2019). The good, the bad and the beauty of 'good enough data.' In A. Daly, S.K. Devitt, & M. Mann (Eds.), *Good data, theory on demand* (pp. 54–76). Amsterdam, The Netherlands: Institute for Network Cultures,
- Higgins, E. (2013, February 25). Weapons from the former Yugoslavia spread through Syria's war. *The New York Times*. Retrieved from <http://atwar.blogs.nytimes.com/2013/02/25/weapons-from-the-former-yugoslavia-spread-through-syrias-war/>
- Kazansky, B., Torres, G., van der Velden, L., Wissenbach, K., & Milan, S. (2019). Data for the Social Good: Toward a Data-Activist Research Agenda. In A. Daly, S.K. Devitt, & M. Mann (Eds.), *Good data, theory on demand* (pp. 244–259). Amsterdam, The Netherlands: Institute for Network Cultures,
- Kelley, J. (2002). Knowledge nirvana: Achieving the competitive advantage through enterprise content management and optimizing team collaboration. Fairfax, VA: Xulon Press.
- Kohli, A. (2018). Data-driven human rights investigations. In O. Hahn & F. Stalph (Eds.), *Digital investigative journalism: Data, visual analytics and innovative methodology in international reporting* (pp. 67–77). London, UK: Palgrave Macmillan.
- Li, R., Lei, K.H., Khadiwala, R., & Chang, K.C.C. (2012). TEDAS: A twitter-based event detection and analysis system. *2012 IEEE 28th International Conference on Data Engineering*, pp. 1273–1276. Anaheim, CA.
- Manoff, M. (2004). Theories of the archive from across the disciplines. *Portal: Libraries and the Academy* 4, 9–25.
- McCracken, K. (2017). Archives as activism. *Archive History*. Retrieved from <https://activehistory.ca/2017/04/archives-as-activism/>
- McDermott, Y., Murray, D., & Koenig, A. (2019). *Digital accountability symposium: Whose stories get told, and by whom? Representativeness in open source human rights investigations*. Retrieved from <http://opiniojuris.org/2019/12/19/digital-accountability-symposium-whose-stories-get-told-and-by-whom-representativeness-in-open-source-human-rights-investigations/>
- McMinn, A.J., Moshfeghi, Y., & Jose, J.M. (2013). Building a large-scale corpus for evaluating event detection on Twitter. *Proceedings of the 22<sup>nd</sup> ACM International Conference on Information and Knowledge Management, CIKM*, 409–418. San Francisco, CA.
- Mozur, P. (2018, October 15). A genocide incited on Facebook, with posts from Myanmar's military. *The New York Times*. Retrieved from <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>

- National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, DC: National Academies Press.
- Owens, K. (2019). Improving the odds: Strengthening the prospects for accountability in the Syrian conflict by regulating the marketplace for information on atrocity crimes. *University of Miami International & Comparative Law Review*, 26(2), 369–436.
- Panagiotou, N., Katakis, I., & Gunopulos, D. (2016). Detecting events in online social networks: Definitions, trends and challenges. In S. Michaelis, N. Piatkowski, & M. Stolpe (Eds.), *Solving large scale learning tasks: Challenges and algorithms - Essays dedicated to Katharina Morik on the occasion of her 60th birthday* (pp. 42–84). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 9580). New York, NY: Springer Verlag. [https://doi.org/10.1007/978-3-319-41706-6\\_2](https://doi.org/10.1007/978-3-319-41706-6_2)
- Papadopoulos, S., Zigkolis, C., & Kompatsiaris, Y. (2011). Cluster-based landmark and event detection for tagged photo collections. *IEEE Multimedia*, 18, 52–63.
- Petrovic, S., Osborne, M., & Lavrenko, V. (2010). Streaming first story detection with application to Twitter. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10)*, pp. 181–189. Association for Computational Linguistics. Stroudsburg, PA.
- Pringle, R. (2017, December 8). *Rampant social media misuse puts future of popular platforms at risk*. Retrieved from <https://www.cbc.ca/news/technology/social-media-facebook-twitter-instagram-society-negative-1.4429146>
- The Prosecutor v. Mahmoud Mustafa Busayf Al-Werfalli. (2017). International Criminal Court. Retrieved from <https://www.icc-cpi.int/CaseInformationSheets/al-werfalliEng.pdf>
- Rieks, D. (2013). Social media networks stripping data from your digital photos. *Library of Congress*. Retrieved from <https://blogs.loc.gov/thesignal/2013/04/social-media-networks-strippingdata-from-your-digital-photos/>
- Schwartz, J. M., & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science*, 2(1–2), 1–19.
- Sterling, J. (2012, March 14). *For Syrian activists, YouTube is a sword and shield*. Retrieved from <https://edition.cnn.com/2012/03/14/world/meast/syria-youtube-uprising/index.html>
- Syrian Archive. (2020a). *Tech advocacy: Amount of content preserved, made unavailable and restored*. Retrieved from <https://syrianarchive.org/en/tech-advocacy>
- Syrian Archive. (2020b). *Tools and methods*. Retrieved from [https://syrianarchive.org/en/tools\\_methods](https://syrianarchive.org/en/tools_methods)



Trachtenberg, A. (2008). Through a glass, darkly: Photography and cultural memory. *Social Research: An International Quarterly*, 75(1), 111–132.

Twitter. (2020). Twitter rules enforcement. Retrieved from <https://transparency.twitter.com/en/twitter-rules-enforcement.html>

von Krogh, G., & Spaeth, S. (2007). The open source software phenomenon: Characteristics that promote research. *Journal of Strategic Information Systems*, 16, 236–253.

Warner, B. (2019). Tech companies are deleting evidence of war crimes. *The Atlantic*. Retrieved from <https://www.theatlantic.com/ideas/archive/2019/05/facebook-algorithms-are-making-it-harder/588931/>

Webb Williams, N. (2015). Protest observation and mass self-communication: Meditations on the Arab Spring. Retrieved from <https://ssrn.com/abstract=2579358> or <http://dx.doi.org/10.2139/ssrn.2579358>

Weng, J., & Lee, B. (2011). Event detection in Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain. Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2767/3299>