DATACTIVE

Public OSINT

What is Open Source in Open Source Investigations?

Working Paper Series
No 1/2021
ISSN: 2666-0733

Lonneke van der Velden
University of Amsterdam

Contact the author at
lonneke@data-activism.net

DATACTIVE can be found at
University of Amsterdam
Department of Media Studies
Turfdraagsterpad 9
1012 XT Amsterdam

Data activism: The politics of big data according to civil society (DATACTIVE) is a research project based at the Department of Media Studies of the University of Amsterdam. It is funded by a Starting Grant of the European Research Council (grant agreement no. 639379), with Stefania Milan as Principal Investigator.

The DATACTIVE Working Paper Series presents results of the DATACTIVE research project. The series aims to disseminate the results of their research to a wider audience. An editorial committee consisting of the DATACTIVE PI and Postdoctoral fellows reviews the quality of the Working Papers. The Series aims to disseminate research results in an accessible manner to a wider audience. All Working Papers are available for download in PDF format at https://data-activism.net.

Readers are encouraged to provide the authors with feedback and/or questions.

Public OSINT: What is Open Source in Open Source Investigations?

Abstract

In the past few years we have seen collectives and action groups emerging that use digital data for the online investigation of events. Groups like BellingCat, Syrian Archive, and Amnesty's Digital Verification Corps are projects that make use of the enormous pool of social media data, satellite images, and leaks that circulate through the digital domain. Practices of online 'Open Source Investigations' (OSINV), sometimes referred to as online Open Source Intelligence (OSINT), form the core subject of this article. The article reflects upon civil society driven projects that conduct such investigations and articulate their aims in terms of the public interest and human rights. In particular, it zooms into how these projects approach the notion of openness in practices of open source investigations by making use of the notion of 'epistemic cultures'. The paper concludes by arguing that if we take openness as a unit of demarcation we see it is context dependent: practitioners define the openness of a source differently. If we take openness as a value we see it is shaping the way investigatory skills are being publicly learned.

Keywords: Open Source Investigations, datafication, data activism, epistemic cultures, transparency.

Introduction: Openness in Online Open Source Investigations

In preparation of this paper, I attended a workshop in Online Open Source Investigations. At a certain moment, the trainer showcased how her organisation had discovered a way to bypass a blockade for a specific form of 'search' on Facebook, an insight they had shared widely via Twitter. One of the attendants in the workshop asked why they would at all want to share this publicly, as Facebook would quickly be able to react with an update. The trainer replied that, rather than keeping their tricks private, sharing their knowledge was more in line with the 'ideals of openness in the OSINT scene', even if the result of sharing would mean that their intervention would be only of temporary use. That comment triggered me to look further into how 'openness' is constructed through the work of people practicing online Open source Investigations. When does 'open' refer to 'openly accessible sources', as prescribed by the standard definition of Open Source Intelligence? And when do we see other meanings of openness? Do perceptions of openness play a different, and possibly, more formative role that impacts also the methodology of investigations?

In this paper, I use the notion of 'epistemic culture' (Knorr Cetina 1999) as a lens to look more systematically at the practices, methods and values that are articulated by open source investigations projects. That means the study focuses on the ways by which practitioners understand what count as 'open' sources, how they define epistemic agency, and what are organizational structures through which they conduct their work (Knorr Cetina and Reichmann 2015). In the analysis of the type of work conducted by these projects, the notion of epistemic culture helps me not only to bring into view objects of knowledge, epistemic agency and organizational formats. The notion of epistemic culture also helps in highlighting a certain dynamic through which the setting of their work – which is predominantly the public domain - folds into methodological choices.

The paper contributes studies into 'data activism' (Milan 2017; Kazansky et al. 2019; Gutiérrez 2018; Milan and Velden 2016; Baack 2008), as it engages with how citizens reshape contemporary data practices through their particular positioning, social values and approaches to technology. It also contributes to an emerging strand of literature which focuses on the intersection of data investigations and human rights in particular (Anden-Papadopoulos 2020; Gray 2019; Dubberley, Koenig, and Murray 2020; Gregory 2019; 2010). What this article tries to add to that body of literature is a reflection on what 'open' means in Open Source Investigations, as it seems that openness is not always clear cut. Rather, it is being constructed through different working styles. This is one of the main outcomes of bringing the notion of epistemic cultures in conversation with Open Source Investigations.

The empirical work rests on one week of participative observation in a workshop, in which I joined a group of people in learning the basic principles and technicalities of doing online open source investigations, five semi-structured interviews, conducted in the context of the DATACTIVE project at the University of Amsterdam, and publicly accessible information about Open Source Investigations (eg. blogs, tutorials, online interviews). The projects referred to in the paper can be considered as (non scientific) laboratories that experiment with online open source investigations. They are potentially indicative of an emerging field of 'public OSINT', a question I will address further at the end of this article.

OSINT or OSINV? On terminology

The subject matter of this article is Online Open Source Investigations, which can be abbreviated by the acronym 'OSINV'. More commonly used is the acronym 'OSINT', which refers to Open Source Intelligence. OSINT refers to intelligence practices on the basis of publicly available sources. It is 'a process whereby police or other investigative agencies gather and analyse data that are in principle accessible to any organisation or individual'(Trottier 2015, 531). The website of the CIA (2017) explained on OSINT saying that there is a range of publicly available material highly relevant to intelligence purposes, in addition to secret information.[1] OSINT, according to them, provides an entry point to 'an endless supply of information that contributes to our understanding of the world' and at a 'relatively low cost' (CIA 2017).

OSINT is also conducted by other parties than intelligence agencies, including private companies, (data) journalists, NGOs, and human rights activists, and civil society actors. The projects discussed in the article do not work on behalf of formal investigative (state) agencies. They don't work for the state nor for OSINT contractors. The people interviewed are, for instance, active at organizations that see their work as beneficial for human rights or of relevance to anti-corruption reporting. They articulate their aims to be in the public interest and have formal ties or casual collaborations with official news outlets, journalistic networks, academic institutions or funders in the field of human rights, journalism, open knowledge and organised civil society. Hence, even though some of the respondents work on a voluntary

[1] They mention: the internet; Traditional mass media; Specialized journals, conference proceedings, and think tank studies; Photo's; Geospatial information

basis, they do receive institutional acknowledgement. Not included in the corpus are groups that are lacking those ties and have been engaged in actions that some have termed digital vigilantism (eg. Anonymous' outing of rapist) (CNN 2013), even though this would be interesting to research as well (cf. Trottier 2017).

One of the reasons why some projects prefer the term online open source 'investigations' instead of 'intelligence' is exactly because the latter carries a connotation with state intelligence, which is a term not all human rights defenders feel comfortable with (Respondent 3). OSINT and OSINV don't necessarily exclude one another, but if we want to dig deeper into the definitions, Dubberley et al. (2020) distinguish the two as follows in one of the newest handbooks on digital witnessing. Whereas in OSINV the use of online open source information is tied to investigative processes (and situated in legal processes with the aim to determine wrongdoing (p. 9), OSINT is:

> 'information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement. While intelligence operations are distinct from criminal investigations, OSINT practices, such as real-time monitoring, may inform certain aspects of open source investigations.' (Dubberley, Koenig, and Murray 2020, 9-10)

From the interviews it turned out that whereas some practitioners want to disassociate from the notion of intelligence in framing their own work, others are not particularly concerned with the exact wording, whether it is Open Source Investigation, Open Source Intelligence, data driven investigation or data journalism. Again others do explicitly refer to an 'OSINT scene', which finds itself publicly connected via Twitter, and internally via Slack channels (source: Workshop). I consider this messiness in naming as an indication that the community that works on and around online open sources is itself still rather incoherent. Definitions such as the one provided by Dubberley et al. might be attempts to demarcate a community of practice or even to professionalize a field by working on a set of shared terms. I will be using the term OSINV most of the time. Even though the abbreviation is less commonly used, it seems to be the most inclusive term. It is inclusive to those who do not see themselves as intelligence practitioners, yet it is not exclusive to those that do associate with intelligence: not many OSINT practitioners will say they are not doing investigations.

The other complicated term in the context of open source investigations, and at the course of this paper, is 'open source'. As becomes clear from standard definitions of OSINT/V (eg. in Wikipedia), Open Source is usually not associated with 'Open Source principles', a term associated with the free software and open source communities. Free and open-source software is software for which the source code is licensed to be open to be reused, modified, and redistributed. It has been developed within communities that are associated with freedom of information and an open internet (Kelty 2008; Coleman 2012). Open Source in OSINT/V refers strictly speaking only to the sources that are investigated: they are 'openly' available. The interviews conducted for this paper, however, suggest that rethinking the connections between Open Source Intelligence, Open Source tools, and Open Source communities can be productive in some cases, as has also in the past been done by Stalder and Hirsch (2002) and Glassman and Kang (2012). That is because ideals of openness do come to play a role in some of the practitioners' approaches, for instance, when they 'open source' their own methods and tools and are committed to collaborative learning.

What are examples of online open source investigations?

In the past couple of years, we have seen quite a few projects in the realm of civil society, journalism and activism, that articulate their work as investigatory practice on the basis of open sources. One of the more famous projects conducting (and probably as well framing) online open source investigations is BellingCat. BellingCat used to be a voluntary collective which has, by now, has turned itself into a company. The founding story behind BellingCat is that the founder, a British accountant named Eliot Higgins, was unemployed for some time during which he turned himself into a weapon analyst by monitoring videos of the Syrian conflict on YouTube, and in that way experimented with a new form of "social media weapon tracking". In 2014, he founded Bellingcat, a collective of investigative journalists and open-source intelligence practitioners (O'Brien 2013). By now, BellingCat has become well known, for instance through their investigation into the downing of the MH17. The Malaysia Airlines Flight 17 was shot in 2014 above Ukraine close to the Russia–Ukraine border. The investigation into the event was led by the Dutch joint investigation team (JIT), which pointed to a Buk-missile being launched by pro-Russian separatists. BellingCat involved itself in the investigation by producing, amongst other things, material related to the Buk-convoy. For that case they managed to retrieve video material regarding the buk convoy, and locate its specific time and place by correlating the content of the video material to, amongst other things, gas prices and images from Google Street View (BellingCat, dossier MH17, 2020).

BellingCat is an example of a project that has started from outside of any formal knowledge organisation. Higgins has been called an 'armchair intelligence practitioner', a term that has been used both in a praising and derogatory way. Like its founder, some analysts that work with BellingCat are self-taught. Others have received training at their previous jobs or studies or at the BellingCat collective itself (Respondent 1). Even though they started as 'citizen investigators', BellingCat became increasingly mentioned as a trusted source in the realm of journalism, and the collective and its members have been awarded journalistic prizes (Jan 2017; European Press Prize 2019). Higgins is now associated with several academic institutions (such King's College London and the University of California, Berkeley (Berkeley Human Rights Center 2020). In 2019, the organisation consists of about 17 staff members, smaller groups of about 20 researchers and a larger network of about 50 contributors.

BellingCat serves as a good example to explain what is OSINT in the public domain. BellingCat self-describes its methods of open source analysis as 'transparent'. Not only the sources should be transparent ('open'); also the process should. The steps in the investigation should be re-traceable as well and published. The group frequently appeals to the public in order to involve people in their investigations, and they offer training.

One could argue that the group is quite unique, considering its history as an amateur collective which has later become a well known and trusted source. However, there exist quite a couple of groups that are developing expertise in piecing together various forms of digital data for investigatory purposes, people working on image analysis, text analysis, satellite image analysis, geolocation analysis and working with data codification schemes. They also use various degrees of open methods and collaborate with volunteers. The existence of multiple projects indicates that BellingCat is part of a larger field of practices (Niezen 2020), which can broaden our perspective on what shape the field of OSINV has. The work by these groups is increasingly being documented. See for instance, Deutch (2020) on Syrian Archive, Gray (2019) on Amnesty's Decoders, and (Banchik 2019) on the Human Rights Center Investigations Lab (Berkeley).

Using epistemic culture as a lens for OSINV

In order to get more insight into knowledge production through OSINV groups, this article builds on the notion of epistemic cultures. The notion was articulated at the end of the nineties within the field of Science & Technology Studies (STS) in order to point at the diversity in modes of knowledge making. The concept directs attention to the 'specific strategies that generate, validate, and communicate scientific accomplishments' (Knorr Cetina and Reichmann 2015, 873):

> 'It brings into focus the content of knowledge-oriented lifeworlds as lived, taken for granted worlds, including specific constructions of the referent (the objects of knowledge), distinctive models of the epistemic subject (scientists), and particular structures of existence with which a field engages.' (Knorr Cetina and Reichmann 2015, 874)

Open Source Intelligence and Open Source Investigations can be approached in a similar way, as a knowledge practice that comes in diverse shapes and forms. Some forms are formally authorised (by, for instance, state institutions); others are conducted by self-organised groups. This study further investigates what sort of knowledge-making is happening in those more alternative niches and hopes to understand what kinds of practices are trusted by those who are active in that field. The way the notion of epistemic cultures is operationalised in this article is therefore not as a concept to be applied to OSINV in evaluative terms (as in: "Is this a specific epistemic culture"?), but rather as a conceptual device which guides the attention of the researcher to be sensitive to certain practices more than others, such as: practices of knowledge validation, the (re)ordering of objects, and organisational formats. This helps depict how these practitioners approach their own work.

Let's first dissect the concept of epistemic cultures to get a more granular understanding of the questions that the concept opens up. For Knorr Cetina and Reichmann, the word 'epistemic' refers to the level of analysis (meaning: truth related goals and practices) (873). The notion of 'culture' refers to the level of practice (with which one needs to gain a 'working familiarity' (874)). More specifically, given the material setting of (most of the) the sciences, focusing on 'culture' would mean bringing into view 'a nexus of lifeworlds and the machineries of knowing that develop within a specialty' (874). Knorr Cetina's work gives several guidelines for doing this. For instance, one question to focus on is: How do scientists construct their epistemic referent, or objects of knowledge? In her work, scientists from different fields had different understandings of material entities, different notions of what constitutes the empirical, and used different vocabularies and metaphors for approaching their instruments. Another guiding question is: How is epistemic agency and authority assigned? In other words: What is the model of the one who knows? For instance, scientists can assign authorship in different ways, for instance, they can attribute authorship to the individual versus the collective research group (875). A third question would be: What are the structures upon which these sciences operate? In her case she looked at the formats that coordinate human and non-human entities. (Here one can think of the scale of the instruments in use, and the kind of cooperation that this brings about.) To get back to the notion of culture: according to Knorr Cetina and Reichman, one can speak of cultural specificities (874) when practitioners make self-referential claims and start policing cultural scripts.

The notion of epistemic cultures has had a whole 'career' itself, having been operationalised to study different research fields ranging from the natural sciences to the humanities (876). The notion has been taken to non-academic, yet knowledge-intensive, contexts as well (such as software uses, public policy, financing) (876). Some have used the notion in a broader way to include social environmental factors (Cole 2013, 38). For example, Simon Cole proposes a social epistemological approach to inquire into the specificities of a (in his case: American) 'forensic culture', which may differ from other science cultures. His paper is of particular interest to this writing about OSINV, because his work also focuses on an 'investigatory culture'. Cole discusses many aspects, including how typical reward structures and audiences play a crucial role in the way forensic knowledge is produced. Hence, in his analysis he includes how the particular audiences for forensic reporting, which are primarily the courts, are generative of forensic culture. He contrasts this to the reward structures such citation indexes and peer assessments, in what we consider as research science (Cole 2013, 40). Forensic scientists have productivity goals in terms of reporting, and they need to speak to the language of a largely (technically) unskilled public. He also points to the fact that American forensic scientists operate in an audit culture 'in the shadow of adversarialism' (40). This working context, in which reports take part in a legal battle in which they will possibly be debunked, has as a practical consequence that forensic scientists are not eager to share their data openly, which in turn leads to a culture with a tendency to 'obscurity' (41). Cole's emphasis on the role of the audience of forensic practice convinced me of the need to discuss the role of the public environment of OSINV investigations in this paper. Because online OSINV culture (in contrast to forensic culture) takes place in a radically different social environment: predominantly online, very public, and to a certain extent it is building on the participation of (known and unknown) publics.

To sum up: In this paper I operationalise the notion of epistemic cultures for the study of a non-scientific phenomenon, for an 'expert culture' outside of science (Knorr Cetina 1999, 246): OSINV projects that work in the public domain. I extend the focused approach presented by Knorr-Cetina with a discussion about the social environment of OSINV projects, triggered by Cole's writings. In my analysis I first make use of three pointers suggested in the work of Knorr Cetina (Knorr Cetina and Reichmann 2015): (a) 'epistemic referent', (b) 'epistemic agency', and (c) 'structures', which I rephrase as 'organizational formats'.[2] Translated to my subject matter, I look at how participants in OSINV projects (a) refer to, or work with, their objects of knowledge; how they (b) stage the actors who know; and (c) what organizational formats make their work possible. In addition, inspired by Cole's analysis of the formative role of audiences and reporting, I take into account how the specific public setting, which is the working domain of these projects, interacts with how their knowledge is (e) validated and (e) communicated.

(a) Epistemic referent: 'open' sources

If we look at what constitutes the epistemic referent of OSINV projects it is the openly available source. Of course do people deal with different topics and issues, depending on what investigation is taking

[2] I use the term organizational formats to clarify that I take these formats or 'ways of organizing' as more fluid than the word 'structures' would suggest.

place. But their object methodologically speaking is the ‘open source’ . However, openness is not for everyone the same, and there are grey zones in understanding what open and closed means in certain contexts. For instance, is a closed Facebook group an open source or not? It is such a group technically open in the sense that you can subscribe to the group, but are there any (social) boundaries to subscribing? Is giving a false name to a Facebook group administrator accepted or not, and does that make the data that are being accessed open sources or not? Taken from the interviews and the field work, some projects advise people to not to do that as it would trespass social and ethical boundaries, while others have a more liberal attitude. On some occasions, security reasons, or concerns about the security of the investigators (sometimes students), are reasons to not dig deeper for certain material, and stop, even though sources are technically open (Respondent 3).

Sometimes open source investigations require ‘special tricks’ to get the information to the surface, as one of the OSINT workshop leaders showcased. One of such tricks mentioned was the sophisticated editing of URLs, such as editing LinkedIn or Facebook URLs. Such subtle modification allows one to query these social media platforms for specific items. That means that sources are not just ‘laying out there’. These kinds of tricks inquire into how the platform is shaped and intervene with how it is (publicly) coded before it can serve (potentially) interesting data.[3]

An interesting explication of the point that what is retrievable depends on your tricks can be found on the project website of ShareLab. They define their work as targeting ‘invisible infrastructures’ and invisible infrastructures are by definition not easy to see. In defining their work, they state: “In most cases, our method does not rely on data available online as a result of open data initiatives or in official documents – it’s rather based on active remote sensing, acquisition of information by using different open network analysis tools or data produced by applications or hardware” (ShareLab 2016). What they bring to the table is the fact that what is ‘visible’ is actually relative to the tools and expertise one has.[4]

That investigators do not always have a ‘rich sea of data floating around’ is shown by a group like Syrian Archive (Syrian Archive n.d.; Deutch 2020), dedicated to document human rights atrocities in Syria. Big online platforms such as YouTube (algorithmically) remove violent material because it violates their terms of service (Respondent 2). This may include material that potentially documents human rights violations. Syrian Archive members try to lobby these corporations to restore this material, in the hope they can include it in their dataset. The dataset is in turn codified in terms of data types, and a list of sources (user accounts/uploaders), through which they try to corroborate material that they have received or collected with information from witnesses on the ground (Respondent 2). That means that part of their work consists of re-opening source material that can subsequently be investigated and turned into witness narratives. Hence, their form of online open source investigations includes restoring ‘lost’ open data through the means of human lobby.

As is described in a recent BellingCat post, finding data sometimes involves paying fees for datasets that are available on black markets: ‘Due to porous data protection measures in Russia, it only takes some creative Googling (or Yandexing) and a few hundred euros worth of cryptocurrency to be fed through an automated payment platform, not much different than Amazon or Lexis Nexis, to acquire telephone records with geolocation data, passenger manifests, and residential data’ (BellingCat, 14 December 2020). Here ‘openness’ is defined against the porosity of data protection.

[3] Some people called this ‘URL-hacking’.

[4] Which in fact also counts for state-conducted OSINT which might be retrievable but is not necessarily public material legally speaking

Finally, many of these labs don't limit themselves to open and online sources in the sense that they try to match open and online data with what they hear from witnesses or other kinds of informants. (Niezen 2020) (Respondents 1,2,3).

In sum: What constitutes 'open sources' are to a certain extent constituted by people's skills, tools, ethical standpoints, definitions of what is closed or protected, and case-specific security concerns, and these differ per lab.

(b) Epistemic agency

The second pointer offered by the notion of epistemic culture is the question: Who is the agent of knowledge production? Several respondents emphasized communal ways of problem solving and distributed expertise (Respondents 1, 2). One needs expertise about language, about geographic regions, geolocation methodologies, scraping youtube, and organisational knowledge (whether this concerns YouTube or Russian data protection measures). Knowledge can be the output from different kinds of data analysis, language skills, local knowledge, and crowd sourced information.

Moreover, beyond the knowledge that exists within the organization, people build on their larger networks, thereby piecing together the various bits of data. This is afforded through the use of social media. For instance, BellingCat presented various cases that were dependent on unique insights from their network (Source: Workshop). For instance, somebody recognising a specific location next to a road on an image, or suggesting that the color of the sand in a picture fits better another area than the one BellingCat's investigators were originally looking at. Or asking botanical experts (BellingCat, 19 March 2019; tweet by Elizabeth Blunt 2019). Or BellingCat making use of locals walking by houses under investigation. One respondent mentioned the lack of hierarchy which allowed the participants to have 'an effect on the environment they are working on' (Respondent 1) , which is also an indication of the distributed way of working.

Prizes are awarded for individual work, still, many of the publications of these groups and collectives appear in the name of the collective.

In sum: The way these groups build on collective contributions, and the way members present themselves and how they assign authorship does not stage a single epistemic subject (Knorr-Cetina) but tends to rather stage 'distributed cognition' (Giere).

(c) Organizational formats

In Knorr Cetina's work she addresses the scale of the instruments that scientific laboratories use which are crucial for their modes of cooperation (Knorr Cetina 1999; Knorr Cetina and Reichmann 2015, 875). The material infrastructure of OSINV labs has the form of a digitised, dispersed, and (mostly) corporate landscape. The instruments that OSINV practitioners depend on are things such as: Google maps, Google Satellite, Invid Plugin, reverse image engineering software, user account verification tools, and all sorts of other tools. The availability of these tools on a global scale allows global cooperation of investigators, peer review, and global outreach (Source: Workshop).

Hence, OSINV investigators depend not only on distributed cognition but also on 'distributed methods' articulated through a whole range of tools (Marres 2012), of which some do and some do not disclose their methodological principles. As Marres argues in the context of social research: 'In the context of digitization (...) social research becomes noticeably a distributed accomplishment: online platforms, users, devices and informational practices actively contribute to the performance of digital social research.' (Marres 2012, 139). From a critical point of view, this means one should pay attention for which methods are being inscribed in OSINV via those tools and what kind of methodology critiques can be articulated (Marres 2012, 152).

This issue – the awareness about the remediation by digital devices - is approached differently per project. Whereas some projects seem to be more liberal about available software as long as it does a good enough job in the investigation (Respondent 3), Syrian Archive states, under tools and methods (Syrian Archive n.d.), that they do not only publish their methods for open source investigations; They also publish their open source software that they use. They only use technology of which the source code is open and reviewable. There are also dedicated tools for open source investigations 'built' with public interest oriented users in mind, see, for instance, Whopostedwhat.com, 'Open for all people who work for public causes' (van Ess, Endresz, and Nemec n.d.).

Finally, to this 'online repository of tools' some investigators add conditioning infrastructures that allow researchers to do their job safely and to prevent leaving too many traces (Respondent 4, 5). This includes the preparation of a secure research profile and network to prevent as much as possible the investigators from being traced themselves. However, for others the awareness of the fact that online investigations also leave traces, does not necessarily translate to such high level security measures. For instance, some recommended photo-apps will be able to use one's data if one uses the free version. Signing up, however, means self-disclosure as well. (source: Workshop).

In sum: The 'machinery of knowing' in OSINV consists of an assemblage of privately owned and publicly available tools. This allows people to collaborate remotely, to operate globally as a network, but also to redo each other's analysis and to compare multiple tools. The active role that the tools play in the way they shape the results and treat personal data is sometimes more and sometimes less critically approached.

The last part of this paper focuses on the public setting in which these people perform their work, as this relatively 'open' stage is used to legitimate and validate their work, and also allows for a broader 'public effect' of these projects, which lies in the area of public learning.

(d) Knowledge validation: sharing methods

When listening to and reading about OSINV projects, publicly sharing the methods and sharing sources seems to be key to how they build trust. This is how BellingCat self-describes in terms of approach and how they intend to build public trust – also with people that might live remotely to where members of the collective are at (Niezen 2020, 165). 'Bellingcat's online reports can be followed by anyone with a modicum of computer skills and the patience to follow the evidence trails step by step' (ibid.). In presenting their work with annotations and their methods as reproducible 'in the manner of a scientific article' (ibid.), BellingCat appeals to scientific ideals in a classic sense. Also other investigators appeal to the ideal of reproducibility. One of the respondents mentioned that by publishing raw data "it becomes a little bit scientific" (Respondent 5).

As already alluded to above, some groups take this further than others. When one of the respondents reflects upon transparency, transparency also refers to about software:

> 'Sure, so may just to add a point to the transparency thing. [The project is] open source … in terms of development and also methodology. Because we are showing the tools that we are using and because we are showing the methods and the workflows, it means that anybody can walk themselves through this process, and that ensures that the findings that we're coming up with can be checked by anybody or challenged by anybody. So we see transparency both in terms of us and the people that are uploading, but also transparency in terms of the wider public, or readership, or other documentation groups. So to anyone really. So that our findings can be transparently evaluated.' (respondent 2)

This emphasis on transparency does not mean that people necessarily believe data is neutral and transparency is a simple guarantee for objective truth. For instance, as one of the interviewees stated: also their list of data categories consists of a selection (for instance it entails a choice for certain 'violation types' recognised by the UN) but they see transparency as a means to be public about this selection also a way of "being transparent about one's biases" (Respondent 2).

In sum: Transparency functions as a token of trust. The way by which public OSINT tries to legitimate knowledge seems to be less institutionally but more technologically en methodologically.

(e) Knowledge communication: public learning

Besides sharing publicly the methods for the sake of transparency, the act of sharing methods also opens up the skills for a larger public. In presenting its aims, BellingCat stresses not only investigations but also teaching and training. BellingCat offers training, including online guides of how-to-do Open Source Investigation and physical (paid) workshops. The online guides include step-by-step explanations, video tutorials and repositories of online tools. Some participants remain connected via Slack. Even though the organisation has professionalised and evolved into a company with an office in The Hague, there remains a group of people contributing to investigations on a voluntary basis.

Like BellingCat, Syrian Archive trains people. Their team consists of about 10 members, 15 volunteers and they work with about 50 students of a number of universities who they have trained and then help them with coding. Other investigators work also with students in dedicated human rights labs in small workgroups.

There is no time for an extended discussion of what one can learn in such workshops, but in the BellingCat workshop the lessons went beyond only technical matters. As mentioned at the start of this article, trainers also highlight why they share knowledge publicly, by referring to the ideals of openness in the 'OSINT scene' (Source: workshop), such as the idea that knowledge should be able to circulate. Hence, these trainings inform its public – which include police officers and security experts of corporations - about open source 'values' that prevail in certain digital (sub)cultures, and hence a particular interpretation of what openness means, in this case: sharing knowledge (with all its risks) is more important than appropriation.

Hence at this point one could sense how OSINV projects can function as experiments in public learning: they first of all bring skills into circulation for evaluating and handling digital data. Parry has discussed how indeed OSINT tutorials can be seen as tools for critical digital pedagogy of interrogating data and its presupposed fakeness or truth (Parry 2019). Moreover, the epistemology articulated through these learning experiments is not only about producing knowledge, but also the sharing knowledge, and the sharing of certain values about what 'openness' should mean. Further more systematic research is needed to see whether these encounters have a critical potential for larger publics.

In sum: Knowledge communication entails moments of public learning, learning of skills and ideals of openness. Hence OSINT itself is being open sourced into a form of 'Public OSINT'.

Conclusion

The notion of epistemic culture in this article has been used as a lens to foreground epistemic practices and legitimations in a range of OSINV labs. What is considered open sources is contextually dependent and is assessed in different ways. Epistemic agency is rather distributed, both in terms of human cognition and contribution, and in terms of material instrumentarium. Practitioners' addressee is the public to which they share knowledge, skills, but also certain values that they regard as important for best practice. Transparency functions as a token of trust. Transparency can refer to the source, the method, and occasionally to the software that is being used. Hence, not only do these actors draw on online methods; when validating their knowledge (in terms of verification and valorisation) some refer to concepts and values that come from social movements that are intimately connected to digital life, such free software (Stalder and Hirsh 2002; Kelty 2008; Glassman and Kang 2012; Kazansky et al. 2019). Hence, if we take open source as a unit of demarcation we see its context-dependency; if we take it as a value we see it is shaping the way skills are being publicly learned. It does also something to our understanding of "OSINT": whereas in the past we would associate OSINT with institutionally embedded work, such as intelligence agencies, OSINT here being opened up, or, more jargonistically, "open sourced" by civil society into a form of public OSINT. And maybe the institutional actors are, in turn, the ones learning from it.

Whether there is a real field or discipline is being formed remains to be investigated. What is currently being shared between a range of actors is self- and community taught expertise (in collecting and analyzing data) in combination with certain principles coming from notions about human rights, free libre and open source thinking, and liberal ideas about open networking. These different standpoints can be recognised in both the diverging approaches to open sources, as well as in how to care for the security of the people involved. Along similar lines, others have argued that open source investigators have no clear way of dealing with the lack of consent of people who upload material online (Banchik 2019).

The coexistence of different standpoints is not surprising given the makeup of the field: journalists, activists, ex-intelligence officers and so called 'armchair intelligence' practitioners, do not necessarily belong to one ‘family of thought’. Knorr Cetina and Reichmann speak of cultural specificities when self-referential claims and the policing of scripts (Knorr Cetina and Reichmann 2015, 874). We do see claims such as ‘the OSINT way of doing’) and, more formally, the publication of standards and protocols that should guide OSINT practices with a human rights angle. Such standards do not only target the specific practices on the level of methodological soundness and ethical concerns, but also on the level of institutions: what do open source investigations need to do to connect to those other, more institutional actors, that evaluate evidence (eg. courts). These attempts are signs of emerging professionalisation and possible ways to form a more coherent community of practice. In that sense, OSINV as a public practice might in fact be an emerging field.

Acknowledgments

This article was researched with the support of a Starting Grant of the European Research Council, awarded to Stefania Milan as Principal Investigator (ERC_StG-2014_639379; https://data-activism.net). The author wishes to thank the members of the DATACTIVE team, for their useful comments and suggestions regarding the many different versions of this paper, and the respondents for their time.

References

Anden-Papadopoulos, Kari. 2020. “Image Activism After the Arab Uprisings| The ‘Image-as-Forensic-Evidence’ Economy in the Post-2011 Syrian Conflict: The Power and Constraints of Contemporary Practices of Video Activism.” International Journal of Communication 14 (0): 34.

Baack, Stefan. 2008. “Civic Tech at MySociety: How the Imagined Affordances of Data Shape Data Activism,” Data Activism, , no. 1. https://krisis.eu/civic-tech-at-mysociety-how-the-imagined-affordances-of-data-shape-data-activism/.

Banchik, Anna Veronica. 2019. Throwing Keywords at the Internet: Emerging Practices and Challenges in Human Rights Open Source Investigations. Dissertation. Austin, TX: University of Texas.

BellingCat. 2019. “Locating The Netherlands’ Most Wanted Criminal By Scrutinising Instagram.” Bellingcat. March 19, 2019. https://www.bellingcat.com/news/uk-and-europe/2019/03/19/locating-the-netherlands-most-wanted-criminal-by-scrutinising-instagram/.

———. 2020. “MH17.” Bellingcat. 2020. https://www.bellingcat.com/tag/mh17/.

Berkeley Human Rights Center, University of California. 2020. “Research Fellows | Human Rights Center.” 2020. https://humanrights.berkeley.edu/about/research-fellows.

CNN, By Todd Leopold. 2013. “Why Anonymous Wants Justice in the Missouri Rape Case.” CNN. October 17, 2013. https://www.cnn.com/2013/10/16/tech/web/anonymous-maryville-rape-case/index.html.

Cole, Simon A. 2013. “Forensic Culture as Epistemic Culture: The Sociology of Forensic Science.” Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, Special Issue: Forensic Cultures, 44 (1): 36–46. https://doi.org/10.1016/j.shpsc.2012.09.003.

Coleman, E. Gabriella. 2012. Coding Freedom: The Ethics and Aesthetics of Hacking. Princeton University Press.

Deutch, Jeff. 2020. “Image Activism After the Arab Uprisings| Challenges in Codifying Events Within Large and Diverse Data Sets of Human Rights Documentation: Memory, Intent, and Bias.” International Journal of Communication 14 (0): 17.

Dubberley, Sam, Alexa Koenig, and Daragh Murray. 2020. Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability. Oxford University Press.

Elizabeth Blunt. 2019. “@henkvaness They Are in Full Bloom Now in the London Area, More Fully out than the One in Your Photo. There Are No Leaves on the Trees in the Background. So, If We Were in London, This Year, I Would Say Late February/Early March.” Tweet. @BluntSpeaking (blog). March 18, 2019. https://twitter.com/BluntSpeaking/status/1107590401274335232.

Ess, Henk van, Daniel Endresz, and Dan Nemec. n.d. “Who Posted What?” Accessed December 24, 2020. https://whopostedwhat.com/.

European Press Prize. 2019. “These Are the Winners of the European Press Prize 2019.” European Press Prize (blog). May 23, 2019. https://www.europeanpressprize.com/winners-european-press-prize-2019/.

Glassman, Michael, and Min Ju Kang. 2012. “Intelligence in the Internet Age: The Emergence and Evolution of Open Source Intelligence (OSINT).” Computers in Human Behavior 28 (2): 673–82. https://doi.org/10.1016/j.chb.2011.11.014.

Gray, Jonathan. 2019. “Data Witnessing: Attending to Injustice with Data in Amnesty International’s Decoders Project.” Information, Communication & Society 0 (0): 1–21. https://doi.org/10.1080/1369118X.2019.1573915.

Gregory, Sam. 2010. “Cameras Everywhere: Ubiquitous Video Documentation of Human Rights, New Forms of Video Advocacy, and Considerations of Safety, Security, Dignity and Consent.” Journal of Human Rights Practice 2 (2): 191–207. https://doi.org/10.1093/jhuman/huq002.

———. 2019. “Cameras Everywhere Revisited: How Digital Technologies and Social Media Aid and Inhibit Human Rights Documentation and Advocacy.” Journal of Human Rights Practice 11 (2): 373–92. https://doi.org/10.1093/jhuman/huz022.

Gutiérrez, Miren. 2018. Data Activism and Social Change. Palgrave Studies in Communication for Social Change. Palgrave MacMillan.

Jan, Benjamin. 2017. “Nederlander krijgt Europese persprijs voor Whatsapp-stuk.” NRC. April 21, 2017. https://www.nrc.nl/nieuws/2017/04/21/innovatie-nederlander-krijgt-europese-persprijs-voor-whatsapp-stuk-8347871-a1555452.

Kazansky, Becky, Guillen Torres, Lonneke Van der Velden, Kersti Ruth Wissenbach, and Milan, Stefania. 2019. “Data for the Social Good: Toward a Data Activist Research Agenda.” In Good Data, edited by Angela Daly, Kate Devitt S., and Monique Mann, 29:244–59. Theory on Demand. Amsterdam: Institute of Network Cultures.

Kelty, Christopher. 2008. “Geeks and Recursive Publics.” In Two Bits: The Cultural Significance of Free Software. Duke University Press.

Knorr Cetina, Karin. 1999. Epistemic Cultures: How the Sciences Make Knowledge. Cambridge: Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674258945.

Knorr Cetina, Karin, and Werner Reichmann. 2015. “Epistemic Cultures A2 - Wright, James D.” In International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 873–80. Oxford: Elsevier. http://www.sciencedirect.com/science/article/pii/B9780080970868104544.

Marres, Noortje. 2012. “The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived.” The Sociological Review 60: 139–165. https://doi.org/10.1111/j.1467-954X.2012.02121.x.

Milan, Stefania. 2017. “Data Activism as the New Frontier of Media Activism.” In Media Activism in the Digital Age, edited by Goubin Yang and Viktor Pickard. Oxon: Routledge.

Milan, Stefania, and Lonneke van der Velden. 2016. “The Alternative Epistemologies of Data Activism.” Digital Culture & Society 2 (2): 57–74. https://doi.org/10.14361/dcs-2016-0205.

Niezen, Ronald. 2020. #HumanRights: The Technologies and Politics of Justice Claims in Action. Stanford, California: Stanford University Press. Pre-published version: https://www.researchgate.net/publication/342962406_Chapter_4_Belling_the_Cat_In_HumanRights_The_Technologies_and_Politics_of_Justice_Claims_in_Practice

Parry, Jason. 2019. Open Source Intelligence as Critical Pedagogy; Or, the Humanities Classroom as Digital Human Rights Lab. In: Interdisciplinary Humanities 36 (1). https://www.academia.edu/42018828/Open_Source_Intelligence_as_Critical_Pedagogy_Or_the_Humanities_Classroom_as_Digital_Human_Rights_Lab

ShareLab. 2016. “About.” Methodology (blog). August 25, 2016. https://labs.rs/en/about/.

Stalder, Felix, and Jesse Hirsh. 2002. “Open Source Intelligence.” First Monday 7 (6). http://128.248.156.56/ojs/index.php/fm/article/view/961.

Syrian Archive. n.d. “Methods and Tools.” Accessed December 24, 2020a. https://syrianarchive.org/en/about/methods-tools.

———. n.d. “Syrian Archive.” Accessed December 24, 2020b. https://syrianarchive.org/.

Toler, Aric. 2020. “Hunting the Hunters: How We Identified Navalny’s FSB Stalkers.” Bellingcat. December 14, 2020. https://www.bellingcat.com/resources/2020/12/14/navalny-fsb-methodology/.

Trottier, Daniel. 2015. “Open Source Intelligence, Social Media and Law Enforcement: Visions, Constraints and Critiques.” European Journal of Cultural Studies 18 (4–5): 530–47. https://doi.org/10.1177/1367549415577396.

———. 2017. “Digital Vigilantism as Weaponisation of Visibility.” Philosophy & Technology 30 (1): 55–72. https://doi.org/10.1007/s13347-016-0216-4.

www.data-activism.net

DAT
ACT
IVE