Responsible Data reflection stories: an overview

Collecting and discussing the unforeseen challenges of using technology and data in advocacy.

Summary

Through the various **Responsible Data Forum events**[1] over the past couple of years, we've heard many anecdotes of responsible data challenges faced by people or organizations. These include potentially harmful data management practices, situations where people have experienced gut feelings that there is potential for harm, or workarounds that people have created to avoid those situations.

But we feel that trading in these "war stories" isn't the most useful way for us to learn from these experiences as a community. Instead, we have worked with our communities to build a set of Reflection Stories: a structured, well-researched knowledge base on the unforeseen challenges and (sometimes) negative consequences of using technology and data for social change.

[1] http://responsibledata.io/events/

We hope that this can offer opportunities for reflection and learning, as well as helping to develop innovative strategies for engaging with technology and data in new and responsible ways.

Methodology

Finding reflection stories

In July 2015, we put out a call on the Responsible Data blog, looking for '**Responsible Data Reflection Stories**"[2], and disseminated it via Twitter and on our Responsible Data mailing list.

Recognising that sharing these stories was perhaps more likely to happen in a more trusted, face-to-face environment, extra promotion of the call for stories was put out on Twitter during events where we were physically present, like the Chaos Communications Congress, the Media Party in Buenos Aires, Argentina, and the Open Government Partnership Summit in Mexico.

Through these efforts, we received a number of email and Twitter tips of leads, and gathered a list of potential leads to follow up on - the large majority from people already within our network.

Going back to the source

Wherever possible, we spoke to at least one – and ideally more – of the people involved. In some cases, though, this wasn't possible–namely, in Stories #4 and #5. In all other cases, we spoke to people involved–ideally from 'both sides' of the story, and they had a chance to review the piece and have input, prior to publishing.

In some cases, we encountered difficulties in reconciling very different perspectives of what happened; thinking about a project from a "responsible data" perspective does inherently lean towards the critical. Because the primary goal of these stories is to highlight **responsible data** challenges of specific cases, they do lean towards giving more space to critical perspectives of said cases, rather than praising them.

Some people we spoke to preferred to remain anonymous, and due to the unique nature of the work they do, this meant that we had to do more than just remove names. Notably, in Story #3, we refer to the user group of an app as a "disproportionately criminalised population", meaning they are an often-discriminated against section of society.

[2] https://responsibledata.io/new-project-responsible-data-reflection-stories/

Internet "chinese whispers"

Our hypothesis when starting this project was that we, as a community, trade in "war stories" that might not be factually accurate. Through following up on the leads we received for this project, we realised that is perhaps an understatement.

Some of the leads for stories were known to us, and others were very well known within certain communities–often repeated as 'fail stories', or as warnings against certain projects.

Here's one, which has been used in the past as an argument **AGAINST** publishing data on where international aid is delivered:

> *"that time that an international aid agency published geolocated data on where their aid workers and facilities were in Pakistan, and the Taliban used this information to attack them directly"*

From a quick online search, a few potentially related events come to light:

- the CIA using a vaccination campaign for hepatitis B to identify Osama bin Laden's hideout
- the Taliban's subsequent attacks on polio vaccination workers in June 2012, when according to National Geographic, Taliban leaders "banned all vaccination programs in the areas under their control"[3] , after which, some international health agencies stopped work, and others continued but under police protection, or while removing their logos from their vehicles.
- Continued attacks, such as when Pakistani Taliban gunmen killed seven Pakistani aid workers in January 2013.[4]

Through all of these related events, though, no relationship seems to be explicit between geolocated data being published and attacks on the aid workers. Of course, that does not rule out the possibility that it **DID** happen–but without any online mention at all of it happening, it seems as though 'chinese whispers' may well have played a role in making this the cautionary tale that it is today.

This indicates that our hypothesis was correct, and cautions us to try and be sure of the source and veracity of a story before spreading it.

[3] http://news.nationalgeographic.com/2015/03/150303-polio-pakistan-islamic-state-refugees-vaccination-health/

[4] http://www.theguardian.com/world/2013/jan/01/gunmen-kill-pakistani-aid-workers

Sector focus

Our original focus was on challenges faced by the use of data in advocacy, but recognising sometimes blurry lines between the two, we have expanded that out to include advocacy and journalism. Data-driven journalism is on the increase, and as Nicolas Kayser-Bril notes, within journalism, "no systematic study of data-driven mistakes has been carried out by academia or professional organizations".[5] While this is far from being a systematic study, we hope it can contribute to our collective intelligence about how just some of these mistakes take place, and the impact they have.

We received a number of tips about responsible data challenges faced by governments, in the private sector, and elsewhere, but for this set of stories at least, we decided to stick to our original mandate. The suggested possible stories, though, does indicate that there is a lot of scope for expansion of reflection stories in the future.

What we learned from the stories

New spaces, new challenges

Moving into new digital spaces is bringing new challenges, and social media is one such space where these challenges are proving very difficult to navigate. This seems to stem from a number of key points:

- organisations with low levels of technical literacy and experience in tech- or data-driven projects, deciding to engage suddenly with a certain tool or technology without realising what this entails. For some, this seems to stem from funders being more willing to support 'innovative' tech projects.
- organisations wishing to engage more with social media while not being aware of more nuanced understandings of public/private spaces online, and how different communities engage with social media. (see story #2)
- unpredictability and different levels of **visibility**: due to how privacy settings on Twitter are currently set, visibility of users can be increased hugely by the actions of others–and once that happens, a user actually has very little agency to change or reverse that. Sadly, being more visible on networks like Twitter disproportionately affects women and minority groups in a negative way. While 'signal boosting' to raise someone's profile might be well-meant, the consequences are hard to predict, and almost impossible to reverse manually. (see story #4)
- consent: related to the above point, "giving consent" can mean many different things when it comes to digital spaces, especially if the person in question has little experience or understanding of using the technology in question (see stories #4 and #5).

[5] http://blog.nkb.fr/data-literacy/

Grey areas of responsible data

In almost all of the cases we looked at, very few decisions were concretely "right" or "wrong". There are many, many grey areas here, which need to be addressed on a case by case basis. In some cases, people involved really did think through their actions, and approached their problems thoughtfully and responsibly–but experienced consequences they had not imagined (see story #8).

Additionally, given the quickly moving nature of the space, challenges can arise that simply would not have been possible at the start.

quickly moving "innovation"

The promise and benefits that technology and increased uses of data can bring, are widely touted. For some advocacy organisations without much experience of engaging with technology in their work, it appears that these benefits are much more clear than the potential risks–which in fact, are much harder to see with low levels of technical literacy.

This promise seems to bring with it an element of speed; using technology generally makes things faster, but thinking through responsible data concerns can slow things down in the short term, while making the projects more successful in the long term. As mentioned above, too–funders seem to be increasingly supporting 'innovative' uses of technology, in some cases without thinking through the responsible data concerns.

As we see in the Mitigation Strategies section below, the time taken to engage with responsible data practices does indeed pay off in the long term–(see story #6 and #3). Those with higher levels of technical literacy–that is, those who are aware of how the technology works, of privacy concerns, more well-versed in ethical debates around what they are doing–are the ones who are engaging more actively in mitigation strategies **DURING** or even **BEFORE** starting projects, rather than afterwards.

COLLABORATION BETWEEN ADVOCACY ORGS AND TECHNICAL PARTNERS

Many of the organisations featured would not consider themselves to be particularly "technical", but in order to develop their project, they partnered with a technical development company. For the most part, effective communication between these two very different types of partners seems to be difficult: they have differing priorities, and very different contextual understandings.

Problems have arisen when concerns of an advocacy organisation, ie. someone with good contextual understanding and experience of working with the community in question, seem to have been under-prioritised or misunderstood by a technical partner.

Mitigation strategies

Using in-person and online networks to ask for help, and being open about the challenges that are being faced.

As outlined in the stories, there are lots of ways in which a project could take an unexpected turn. With the heavy caveat that there isn't **one solution** to solve any of these problems, and that each of these need to be thought through in the context of the project itself, here are some of the strategies that those who are aware of how the technology works and privacy concerns, and who are better versed engaged in while trying to work through the challenges they faced.

CO-DESIGN, COLLABORATION

In terms of community management, **co-design** can be understood as one way to get buy-in and ownership from the community. But building with[6] a community also brings benefits in terms of understanding what responsible data challenges are actually being faced.

Hearing from the community who will be involved in the project (eg. the users of the tool, the people who will feature in the dataset) early on, and repeatedly, seems to vastly increase the likelihood that red flags will be noticed early on, and gives the project owners the chance to redirect course.

BEING FLEXIBLE

For grant-funded organisations, this depends in large part upon the flexibility and understanding of funders. Being able to change course halfway through a project, and thus redirecting funds from one planned use to another, requires either un-earmarked grant money (which many non-profits do not have)–or, funders who recognise the importance of flexibility to build effective projects.

[6] see **http://www.buildwith.org/,** a 'resource hub for direct partnership in governance and technology'

MOBILISING NETWORKS

Using in-person and online networks to gather information from experts is a good way to reach out to trusted people, without having to do anything as drastic as put job advertisements up, or write on an organisation's blog. In many of these cases, this kind of networked advice seems to be offered for free, with the potential of turning into consulting further down the line. To get that first level of advice for free, being able to tap into a network (such as the Responsible Data community) for help seems to really help.

In these cases, abstracting out the different types of expertise needed to address the challenge–for example, rather than asking for a **DIGITAL SECURITY** expert, specifying that an expert is needed to work on, for example, secure storage of images. Additionally, a good strategy seems to be getting multiple opinions from those with more expertise in these areas, rather than relying on one or two. It is likely that people will disagree about what the best course of action is, especially if they do not have the same level of contextual understanding- but gathering their opinions and advice will make it easier to make a well-informed decision.

OWNING MISTAKES

In cases where there have been challenges that were not mitigated against, offering unreserved apologies and owning that a mistake took place is a good first action. Of course, it does not undo any harm that might have been done; but in those cases, getting support from the affected communities will be a necessary strategy in the broader mitigation plan.

ADMITTING UNCERTAINTY

Admitting that there are responsible data challenges shouldn't be seen as a deterrent to the project getting support, but rather the opposite.

In a way, this strategy requires people with a solid understanding of what data can, and can't do. Being able to openly admit that despite thorough analysis of a dataset, or deep expertise in a certain topic, there are multiple levels of uncertainty inherent within the project, is something that not everyone will feel comfortable doing. But in order for the data to genuinely inform positive social change, its limitations need to be explicitly made–not just internally within a team, but also externally, to the audience or community affected.

Conclusion

Despite the very varying settings of the stories collected, the shared mitigation strategies indicate that there are indeed a few key principles that can be kept in mind throughout the development of a new tech- or data-driven project.

The most stark of these–and one key aspect that is underlying many of these challenges–is a fundamental lack of technical literacy among advocacy organisations. This affects the way they interact with technical partners, the decisions they make around the project, the level to which they can have meaningful input, and more. Perhaps more crucially, it also affects the ability to know what to ask for help about–ie, to 'know the unknowns'.

Building an organisation's technical literacy might not mean being able to answer all technical questions in-house, but rather knowing what to ask and what to expect in an answer, from others. For advocacy organisations who don't (yet) have this, it becomes all too easy to outsource not just the actual technical work but the contextual decisions too, which should be a collaborative process, benefiting from both sets of expertise.

There seems to be a lot of scope to expand this set of stories both in terms of collecting more from other advocacy organisations, and into other sectors, too. Ultimately, we hope that sharing our collective intelligence around lessons learned from responsible data challenges faced in projects will contribute to a greater understanding for all of us.

the engine room

This publication is part series found at https://responsibledata.io, produced by the engine room's Responsible Data Program, 2016.

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Access to treatment for HIV/AIDS patients

Challenges faced when trying to improve public health systems to help HIV/AIDS patients in Buenos Aires receive better treatment.

Context

Fundacion Huesped[1] works in Buenos Aires on issues relating to people accessing treatment for HIV, HIV/AIDS and related illnesses as Hepatitis, Tuberculosis and more. In 2015, they set up an Innovation Lab, recognising the growing relationship between health and technology. It is not the main focus of the organisation though, and it is a very new initiative within the organisation.

[1] **Fundacion Huesped**, a non-profit organisation based in Buenos Aires, Argentina was started 26 years ago by a group of doctors working in a public hospital in Buenos Aires. They work in Public Health, specifically on infectology, from a human rights perspective.

The problem *Although receiving treatment for HIV/AIDS is paid for by the government, research has shown that almost half of the people living with HIV/AIDS are not receiving treatment. As part of the work of the Innovation Lab, they wanted to look at the role that technology could play in improving these processes, and making the system easier for patients.*

After talking to patients, doctors and those treating people living with HIV/AIDS, they identified a number of key problems. They talk about this in terms of a '**care cascade**'.[2] For people to successfully receive ongoing treatment for HIV/AIDS, a hypothetical patient would need to go through multiple steps, realising they have the virus, knowing what to do next, visiting a doctor and getting tests done, understanding what their status is, and–most importantly–committing to lifelong treatment.

HIV/AIDS: an intelligent virus *HIV/AIDS is an intelligent virus; if a specific form of treatment is stopped, the virus will come back immune to that specific treatment, so the treatment then needs to be changed for the next course of action. This means that even if symptoms have stopped, patients need to keep taking their treatment, which can be difficult to ensure.*

Challenge: patients getting lost within the public health system

They identified that a major problem is that people get "lost" within the Public Health system–starting treatment, but then not showing up for their next appointment, for example. So, they decided to look at interventions around supporting people as they go through their treatment, looking at potential points at which they drop off or stop their treatment.

"LOST" WITHIN THE PUBLIC HEALTH SYSTEM

To start with, they began looking at the appointment system. Currently, patients have to go in person to make an appointment, between the 1st and the 3rd day of every month, and appointments are only made within the following two months. This is a difficult task for many people; going during certain days, and there are just **three days per month** when appointments can be made. They used to be able to make appointments within a six month window, but realised that too many patients were simply not turning up.

[2] https://www.aids.gov/federal-resources/policies/care-continuum/

REMINDERS

They identified a specific way to mitigate this: sending reminders to patients that their appointment was coming up, to make sure that appointments were being used, and people were getting the treatment they were scheduled to have.

HOW TO GET 'REMINDERS' TO PATIENTS, WITHOUT VIOLATING THEIR PRIVACY?

Patients at this hospital fit into quite a specific demographic; they are attending public hospitals, which means they are likely to be from a low-income background– otherwise they would go to the better resourced, and more expensive, private hospitals. This means they are unlikely to have smartphones, but often feature phones or access to shared landlines, and are big users of Facebook. Many potential users say they "don't have an email, but they have a Facebook."

Fundacion Huesped went to a hackathon and developed a calendar tool sending reminders via SMS and email, and potentially Interactive Voice Response (IVR) as well. Their aim with this tool is to remind the patients of their appointments so that they don't miss them, and decrease the rate of unattended appointments; so naturally, they want to get reminders to the patients using communications channels that are already used by their target demographic.

But most, if not all, of the most commonly used communications channels are largely insecure. Without knowing who has access to those communications channels, how can they be sure that only the person in question will get the message, and not be read or heard by someone else? In the case of shared landlines, they thought about leaving cryptic messages that would mean something to the patient, but be incomprehensible to others who might hear it.

National AIDS Law *In Argentina, there is a National AIDS law that prohibiting anybody from identifying HIV/AIDS patients by their name or DNI (national identity number.)*

With more than half of the people living with HIV/AIDS not receiving treatment, it's clear that there are problems here; clearly, there is a lack of adequate systems in place, such as a digitised appointments system, or reminders to those with appointments. Though Fundacion Huesped have done a lot of work towards identifying concrete problems, and working towards solutions, actually implementing the solutions in a responsible way is proving difficult.

NGOs AND THE GOVERNMENT

As an organisation, they are faced also with resource and time constraints; they are an NGO, not the government, so they are subject to any changes that government might make. The tech solution that they have come up with is dependent upon the current system in place, so if the government choose to change this system, their work would be totally useless.

Mitigation strategy

Although they have now developed a tool to address some of the issues associated with making an appointment and receiving reminders to attend the appointment, they are not rolling it out yet. They are very cognisant of the risks here–both in terms of the patients as individuals, and for them as an organisation, due to the National AIDS law that prohibits anyone from identifying people living with HIV/AIDS.

What next?

They have met with the Ministry of Health to discuss this law, and what this means for their potential technology solutions. The Ministry of Health encouraged them to go ahead with the project, as “not everybody looking for an appointment at the department was living with HIV/AIDS”–so, essentially encouraging them to base their project on a potential legal loophole.

Conflicting messages *They noted, though, that the Ministry of Health themselves have an electronic appointment system that they use in other types of department, but not in Infectology Departments. This could imply that they don’t want to rely on that loophole themselves.*

They are gathering together a group of IT Security experts to work on an architectural solution for this problem, looking at where the connections are between the systems in the hospital, and which security layers need to be in place to be sure that there are no security breaches.

They are also speaking to **private hospitals**[3] to work out how they deal with these issues, and see if they can gain some insights from the technology solution that they already have in place. Their strategy at the moment is to gather experience and expertise from others, and try and come up with a solution together.

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room’s Responsible Data Program, 2016.

[3] **www.hospitalitaliano.org.ar/infomed/**

Mental health-related alerts from Twitter

The case of the Samaritans[1].

Context

Samaritans' work happens primarily through people in need contacting them, via email, telephone, or in person, in addition to their outreach work including workplace training, and work in schools. Samaritans launched the first 24-hour telephone helpline in the UK, and they train volunteers to be "listeners", helping people work out their own way forward.

They identified a number of cases of individuals sharing their thoughts about suicide on Twitter, not receiving a response and tragically going on to take their own life; so, they started a new project to try and prevent this from happening.

[1] **Samaritans** are a non profit organisation based in the UK, with 201 branches across the UK and the Republic of Ireland, and a volunteer community of approximately 20,000 people. They provide emotional support to people in times of need.

Their aim:

moving the Samaritans' "listening ethos" into the realm of social media.

On 29 October 2014, Samaritans launched the Samaritans Radar, an "online app designed to offer people a second chance to see a tweet from someone they know who might be struggling to cope."

The app, Samaritans Radar, monitored the twitter feed of app users to see if anybody followed by the user had tweeted specific keywords or phrases that had been identified as being more used by people who are struggling to cope, such as "I hate myself". If a tweet was found with those keywords, the user would receive an email with a link to that tweet, along with suggested guidance of actions they could carry out.

The response to the application was mixed; some lauded its innovative approach to using social media, but others–notably, many from the mental health community in the UK–reacted strongly against the app.

In response to the negative reaction to Radar, the Samaritans suspended the app just nine days after the launch, and in March 2015, they announced that the app would be permanently closed, and all associated data would be deleted.[2]

How it worked

By default, searches for keywords/phrases were carried out on the tweets of everyone followed by users who signed up to the app. The keywords and phrases were based on research undertaken by Jonathan Scourfield at Cardiff University as part of the **COSMOS project**,[3] looking at **the relationship between social media and suicide**.[4]

If a certain user signed up, then all of the tweets from anybody that they followed were then included within the app and scanned for the designated keywords; and based on the use of certain words which they had identified as commonly being used to indicate "suicidal intent", the tweets were included in notifications to the app user.

As a result, a person's tweets could have been included within the app without that person having any idea, and people following them could have been notified if the Twitter user used a predefined "trigger phrase" unknowingly. At the start the Samaritans site made this very obvious, saying that the people a user follows on Twitter will not be notified that a user has signed up to the app, and all alerts would be sent directly to the user's email address. The Samaritans justified this decision by saying:

> *The app works in such a way that the alerts sent out are only seen by the subscriber, who would have sight of the information anyway. Samaritans does not monitor the tweets or view them – we're just giving people who have*

[2] http://www.samaritans.org/how-we-can-help-you/supporting-someone-online/samaritans-radar#10mar
[3] https://www.cs.cf.ac.uk/cosmos/
[4] https://www.cs.cf.ac.uk/cosmos/research-on-social-media-and-suicide/

> *signed up to Radar a second chance to see a call for help, which they might have initially missed, from a friend that is in need of support."*
> *Sophie Borromeo, director of communications at the Samaritans,*
> **quoted in the Guardian on Nov 3rd 2014**[5]

Essentially, what the app did was make very visible and explicit anything that someone tweeted publicly with predefined "keywords". Technically, anything tweeted from an open account (ie. as opposed to a "closed" account where the user approves individually anyone who wants to follow their tweets) is indeed public, but often it is thought of by users as 'private', especially by users who have very few followers, or who envision that very few people see their tweets.

What's different between this and how Twitter usually works?

Given the fleeting nature of Twitter–that what you see changes rapidly, that it's easy to miss tweets that are sent out by people at certain times of day, tweets are understood to have different levels of visibility. If you tweet at another person without putting a "." before it, for example, only you, that person, and people who follow both of you will see it in their timeline. So, it's a reasonable assumption for a Twitter user with few followers, tweeting when the majority of those followers are asleep, that very few (if any) people will actually see that tweet. What the app essentially did was change that level of visibility.

Mixed responses

Changing that level of visibility, and flagging up every occasion when someone tweeted with those keywords was not seen as helpful by some members of the mental health community.

As written by a former Samaritans volunteer,

> *"How likely are you to tweet about your mental health problems if you know some of your followers would be alerted every time you did? Do you know all your followers? Personally? Are they all friends? What if your stalker was a follower? How would you feel knowing your every 3am mental health crisis tweet was being flagged to people who really don't have your best interests at heart, to put it mildly? In this respect, this app is dangerous. It is terrifying to think that anyone can monitor your tweets, especially the ones that disclose you may be very vulnerable at that time"*–**blog post, Oct 29, 2014**[6]
> *by @elphiemcdork*

[5] www.theguardian.com/voluntary-sector-network/2014/nov/03/samaritans-radar-twitter-mission-charity

[6] emsyblog.wordpress.com/2014/10/29/the-samaritans-radar-app-the-problem-is-right-there-in-the-name/

But others working in the charity sector praised the Samaritans for leading the way in moving their work into the digital realm, **saying that controversy was inevitable with such innovation**.[7]

Challenges faced

As highlighted by those within the mental health community themselves, people suffering from mental health issues felt like the app meant they might have had to self-censor on Twitter. Some felt like their privacy was invaded by the app changing that level of visibility; essentially meaning that they couldn't casually tweet something without others being notified.

For those facing online or offline violence, the app provided potentially more information to allow stalkers or bullies to target people when they were at their most vulnerable. By making it easier for those concerned to see "worrying" tweets, it also made it easier for others–perhaps those with malicious intent–to see the same tweets, and thus identify times of particular vulnerability.

It seems as though during development of the app, primarily positive and useful uses of the data and the app itself were envisioned. Additionally, it was assumed that writing anything negative on Twitter was an explicit cry for attention and cry for help, rather than simply an expression online of one's feelings at that time. As it turned out, many people in the mental health community use Twitter as a way of expressing themselves in a more intimate way than had been envisioned given the technically "public" nature of tweets.

Mitigation strategy

In terms of the technical functionalities, the app included a "whitelist" function; initially, if organisations didn't want their tweets to ever be included within the tweets that were scanned by the app, they could send a direct message to @Samaritans. This was designed for organisations who tweeted regularly with the designated "trigger phrases". Following the negative feedback, this functionality was extended to individuals who didn't want their tweets to be included; however, by default the app was 'opt-out' rather than 'opt-in', meaning that users had to write to the Samaritans to get on the whitelist.

The app was suspended nine days after launch, and permanently closed five months after it began. During the nine days after launch, the Samaritans team put out a number of updates and press releases, all of which are still public and available online.

[7] **www.theguardian.com/voluntary-sector-network/2014/nov/03/samaritans-radar-twitter-mission-charity**

Following the app's suspension, the organisation issued an apology which stated:

> *"We've learned that we must consult even more widely than we have done in the development of Samaritans Radar and we will continue to respect and better understand the diversity of existing communities and users. To this end, we will be holding a series of consultation events as well as continuing to gather views via an online survey from as wide a range of people as possible."*

Importantly, following the app's closure, Samaritans carried out a 6-month knowledge and learning project entitled Digital Futures, with the stated aim of "finding out people's views on the opportunities and challenges for emotional support and suicide reduction presented by the online environment." Their comprehensive findings were all published online in November 2015, and made very clear that they had carried out a broad public consultation, talking to privacy and data experts. These findings outlined areas that the Samaritans could improve in for future digital engagement, suggested by their users, and showed clearly that they had learned from their experience with Radar.

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room's Responsible Data Program, 2016.

Responsible Data Reflection Stories 3

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Creating an app for vulnerable communities

Using technology in order to reduce the incidence of violence among criminalised citizens.

Context

The primary organisation works to support and reduce violence against a disproportionately criminalised population–that is, a group who face disproportionate violence and social exclusion, and who are often treated as criminals without reason. For reasons of anonymity, the group they work with will be referred to throughout this case study as a criminalised population.

THEIR AIM: USING TECHNOLOGY TO EMPOWER MEMBERS OF THE CRIMINALISED POPULATION TO SHARE KNOWLEDGE AMONG THEMSELVES, ULTIMATELY REDUCING INCIDENCES OF VIOLENCE.

The primary organisation are working together with a social enterprise technology company to develop an application that members of the criminalised population can use to report incidents of violence. Their report is then sent out to members of the same community, who are in a similar geographic location to where the report originated from, and, if consent is given, the report is also shared with the police, with no details given about who reported it.

This kind of scheme has been happening in an analogue way for a relatively long time, but this is one of the first attempts to bring it into the digital space. It aims to help inform this specific community through **COLLECTIVE INTELLIGENCE**–helping others to inform their peers to keep them safe.

How it works

Currently, they use a system whereby members of the criminalised population can input reports anonymously via their website, then employees of the primary organisation are tasked with summarising these into shorter reports that can be sent out to other affected parties. With the app, they are exploring a new way of sharing this information as a peer to peer service, thus getting rid of the need for summarising and moderating to be done by the primary organisation.

Their partner company is managing the technical requirements of the app, so the primary organisation does not have direct access to the data. Because this is a partnership between two different organisations, with relatively different aims, they are working hard on negotiating an agreement between the two parties that meets both of their expectations and requirements.

Only members of the primary organisation's network have access to the app; and to become a member, they only need to submit their username and email address. This level of "membership" is kept deliberately low to make it as easy as possible for members of the criminalised population to sign up–by not having to put in names, they want to make it easy for them to remain anonymous with the app. They then fill in specified fields through the app to submit their report.

If the person submitting the report gives consent, the report is shared **ANONYMOUSLY** with the local police force, but without giving any further details about the person who submitted the report.

Challenges faced, and how they're being approached

Personally identifiable information

Under UK law, the reports which are sent out **CANNOT** contain personal information about the perpetrator, as at that point they are alleged to have done a crime, but have not yet been proven guilty. There is also a risk that the perpetrator might find out that they have been reported, leaving the person who reported the crime in potential danger. Balancing the reports sent out to be informative enough so that members of the criminalised population in the same area as where the crime was alleged to be committed can identify and avoid dangerous situations is a main challenge.

Planning for future situations

The Primary Organisation is mindful of the risk that potentially, the police or the UK justice system could issue a court order and get access to their data. With that in mind, they are actively practising **data minimisation** to make sure they have the minimum amount of data required.

They have also found that members of the criminalised population are more likely to be deterred from registering if they have to give lots of personal information, so data minimisation as a principle has multiple benefits. Those who are sending the reports do not want to give any information that might potentially be passed on to the police about their places of work, or any other details about their work.

Working in partnership

As the app is the result of a partnership between one topical focused charity organisation, and a tech-focused social enterprise, their aims have been somewhat different during the development. For the Primary Organisation, their main concern is making sure that **no harm** comes to any of the members of the criminalised population. The social enterprise, however, communicates in terms of "percentage risk"; whereas **any percentage risk at all is too much for the Primary Organisation**.

As the social enterprise is more focused on innovative tech solutions, they are keen to develop new tech solutions–this isn't an aim of the Primary Organisation though, who simply wants to focus on empowering members of the criminalised population to stay safe. Balancing between these different priorities has been a challenge, but they are both working with legal experts to make sure that they have clarity over important points in their partnership–such as who 'owns' the data, especially in case of one of the parties ceasing to operate.

The Primary Organisation does not have the in-house tech capacity to manage or develop the app, which is why their partnership is especially useful. But this has difficulties in terms of introducing dependency from the social issue-focused organisation, to the tech-focused social enterprise.

Moderation of content

Up until development and roll out of the application, when the system used was a website, these reports have been written by employees of the Primary Organisation, all of whom have undergone substantial legal training to make sure that they don't release any reports that could have potential legal consequences. However, with the application, a mode of peer-to-peer sharing is being explored, which means that the reports might go out without anyone from the Primary Organisation seeing them.

To mitigate against this, they have put a number of safety guards in place within the 'report' function in the app to ensure that potentially litigious information cannot be put in; for example, no names of perpetrators can be submitted. They are conscious, though, that should an app user actively try to circumvent these safety guards, it would likely be possible to do so.

They are also exploring how effective these safety guards are in practice, through a pilot phase roll out, and they are mindful of the fact that perhaps this methodology simply won't work. Suggested alternatives to the automatic peer-to-peer report sharing would be reports going first to a moderator who "approves" them before they appear on the app; or a functionality where a moderator can quickly delete the report across all devices, if it proves to be unsuitable for sharing.

LISTENING TO THEIR COMMUNITY

The application was developed through a co-design process, working with members of the criminalised population to figure out the most effective and useful tool for them. The Primary Organisation is proceeding slowly and thoughtfully with the application, trying it out in small areas first, and very clearly putting the focus on their safety throughout, showing willingness to pull functionalities if they could put their community at risk.

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room's Responsible Data Program, 2016.

Responsible Data Reflection Stories 4

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Amplifying narratives from social media

Using mainstream media platforms to pick up on important conversations.

Context

Buzzfeed, “the social news and entertainment company” aim to provide “shareable” content online to their global audience of, as of November 2015, more than 200 million people around the world.

Reporting from Twitter

In mid-2014, a conversation started on Twitter about an important topic; sexual assault. But this conversation was different to many others on the same topic- it was asking survivors of sexual assault what they were wearing when they were attacked. A Buzzfeed journalist noticed the thread, and, seeing a new angle on a crucial topic, decided to write an article about it.

She tweeted at certain people who had responded to the thread, and asked them if she could use their tweets in a Buzzfeed post, offering in the initial tweet to blur their names or their photos.

She **posted the article**[1] with the tweets and blurred photos as agreed upon with the individual tweeters: but after the article was posted, there were mixed reactions to it online. Some were upset as they didn't realise that she had actually asked the individuals quoted for permission, and instead thought she was co-opting a "private" conversation, for the sake of a story.

The person who asked the initial question was angry that she had been included in the story without her specific consent–as only people who **RESPONDED** to the thread had been contacted. But others in the same thread were grateful to her for picking up on it and amplifying important narratives shaped by survivors.

Perhaps more tellingly, in response to this case and the subsequent backlash, a number of other comment pieces were written looking at the ethical situation of using embedded tweets of such a personal nature.

These revealed vast differences in the way that journalists and big news outlets think about and use other people's social media data; some of the opinion that asking for any sort of permission was unnecessary, others pointing out the potential harms.

The challenge: what is "consent" when it comes to using someone's tweets in a news article?

Similar to Reflection Story #2, a lot of this boils down to **LEVELS OF VISIBILITY**. Even though those who were directly quoted in the article were contacted in advance via Twitter, they may not have realised just how popular the post would become, and thus **HOW VISIBLE THEIR TWEETS WOULD BECOME**.

Essentially, though they might have agreed to their tweet being used, it's reasonable to expect that they had no idea what might happen next. In this case, the article was extremely popular, so both the article and the tweets within it got a lot of attention.

It's a sad truth that trolls on social media are common, especially around issues that are particularly important to women or marginalised communities. With this in mind, visibility can have major consequences, such as online violence against women.

Different understandings of "public" and "private" conversations make this situation more complicated; people who are replying to a tweet (especially one about such an intimate topic) may legitimately expect that very few people will see their tweet–especially if they make the active decision to **WRITE IT WITHOUT A . BEFORE THE OTHER PERSON'S HANDLE**.

[1] www.buzzfeed.com/jtes/sexual-assault-survivors-answer-the-question-what-were-you-w#.eb0MnnDZBA

Replying to people on Twitter

If a Twitter user replies directly to another in a thread, usually that reply will only appear to that user, and ANYBODY WHO FOLLOWS BOTH OF THOSE PEOPLE. *To make that reply appear as a usual tweet–ie. in the timelines of anyone who follows the person writing–is by adding a '.' before the @handle.*

Given how tweets work, embedding one tweet in an article could potentially lead readers to other tweets in the thread with just one click, which could increase their levels of visibility, too–and they would not have been aware that this might happen.

The case was a valuable example in journalistic ethics, and the reactions to it revealed how opinions among other journalists differed greatly. There does not (yet) seem to be an "industry standard" about the ethics of embedding tweets on sensitive topics, though there are some especially nuanced takes on the issue, such as **this one from The Cut**,[2] which aptly describes the situation:

> *This debate seems symbolic of the growing tension between news media and social media within feminism.. For journalists, these [situations] require an ethical axis beyond public-private — one that acknowledges* **the high personal stakes**[3] *these conversations involve for their participants.*

What could have happened differently?

The trickiest thing about this example is that the journalist in question behaved just as thoughtfully and ethically as she could have done; she asked permission clearly from each individual before embedding their tweet, and offered to remove names and photos. She was upfront about writing for Buzzfeed, and where requested, she sent links of the published article back to the people involved–something that the vast majority of journalists don't do. There is also now a correction on the article saying that more photos have been blurred, and tweets removed upon request.

To look at what others have done, in another piece talking about this case, the tweets were quoted without attribution, with the following disclaimer:

> [EDITOR'S NOTE: *These replies appear without attribution to protect the privacy of users who did not anticipate that they would be quoted.]*
> *-from* **The Root**[4]

[2] https://nymag.com/thecut/2014/03/twitter-rape-and-privacy-on-social-media.html

[3] https://twitter.com/theferocity/status/444234154495213568

[4] https://www.theroot.com/blogs/the_grapevine/2014/03/sexual_assault_and_women_s_attire_twitter_stories_defy_myths.html

Removing attribution entirely is one way of getting the content into the piece, but then, of course, removing the person from the story could also be problematic. In fact, the empowering angle of highlighting women's stories, and giving them space to share their own stories, could be dampened slightly with this method.

Public/private/something in the middle

From the posts, tweets, and commentary articles on this case, it's clear that **PEOPLE HAVE DIFFERENT EXPECTATIONS OF PRIVACY** despite Twitter being a public platform.

That the journalist in question did actively engage with a number of mitigation strategies to avoid harm, and yet still faced such backlash afterwards, highlights the difficulties of this reflection story. It wasn't the first, and it won't be the last example of this tension between amplifying important stories, versus putting the spotlight on certain people and increasing their visibility.

Verification of social media

Fact-checking and clarity is crucial when communicating on social media: the case of UNHCR on Twitter.

Context

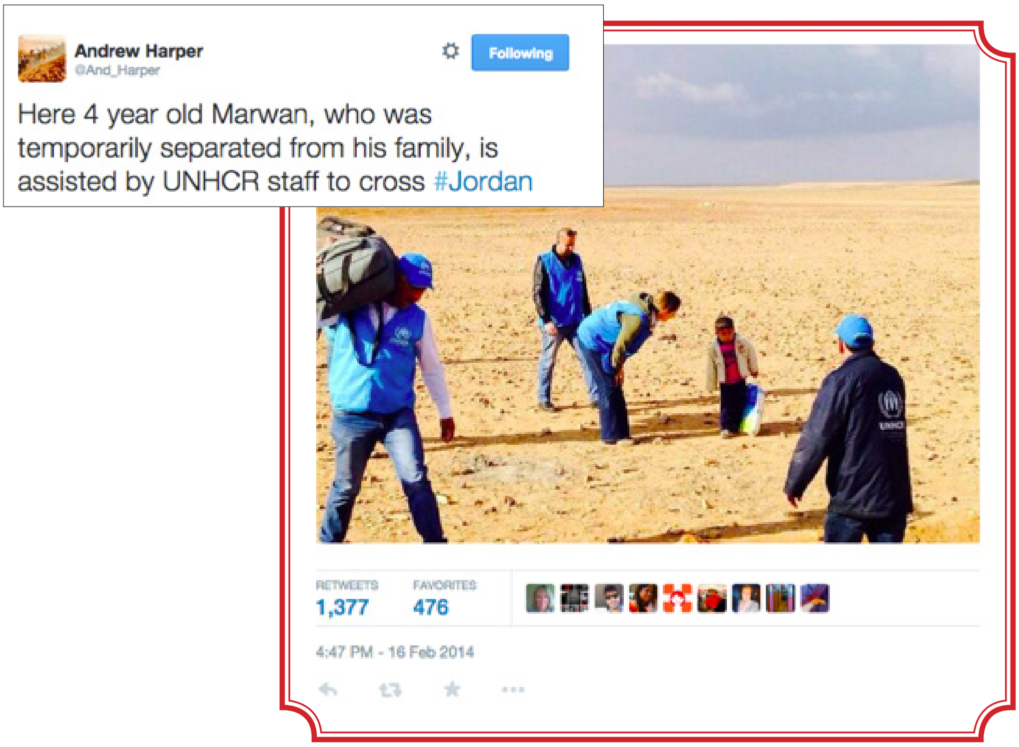
The Office of the United Nations High Commissioner for Refugees (UNHCR) is mandated to lead and co-ordinate international action to protect refugees and resolve refugee problems worldwide. They are very active on Twitter at the handle **@Refugees**;[1] as of December 2015, they have 1.88 million followers, and they have tweeted over 27,000 times. Communicating to the public about what they're doing is important for a number of reasons; for political reasons, as well as garnering public support through donation campaigns.

Story outline

In February 2014, UNHCR representative in Jordan, Andrew Harper tweeted this photo, showing a 4-year-old, Marwan (not his real name), who was "temporarily separated from his family."

It was a heartwarming photo, showing the need and results of UNHCR workers coming to help the boy. It got picked up by relatively mainstream media too–

[1] https://twitter.com/refugees



Andrew Harper
@And_Harper

Here 4 year old Marwan, who was temporarily separated from his family, is assisted by UNHCR staff to cross #Jordan

RETWEETS 1,377 FAVORITES 476

4:47 PM - 16 Feb 2014

"**Refugee named Marwan, 4, found wandering the desert alone**",[2] said the New York Daily News, and **Bustle used it as an example**[3] of how children have especially been suffering from the Syrian conflict.

The original photo from Andrew Harper's account got retweeted over 1,300 times, gaining more coverage after CNN anchor Hala Gorani retweeted it, and so the chinese whispers began–in Gorani's interpretation, Marwan was "crossing the desert alone", and this tweet (now deleted) was retweeted almost 10,000 times, according to the **Columbia Journalism Review**.[4]

But following fact-checking on the photo and tweet by the Guardian, it came to light that the situation described was more complex than it first appeared. According to a UNHCR press officer who was at the border -

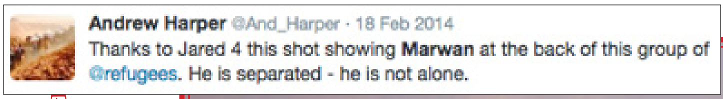
> *"Let me say first, the child was temporarily separated. He was a tiny bit behind his family. His family were ahead and he was just straggling behind. That's the story. He didn't enter as an unaccompanied minor ... he was literally 20 steps behind,"*-**The Guardian, Tues 18 Feb 2014**[5]

[2] www.nydailynews.com/news/world/refugee-marwan-4-found-desert-fleeing-war-syria-article-1.1617606

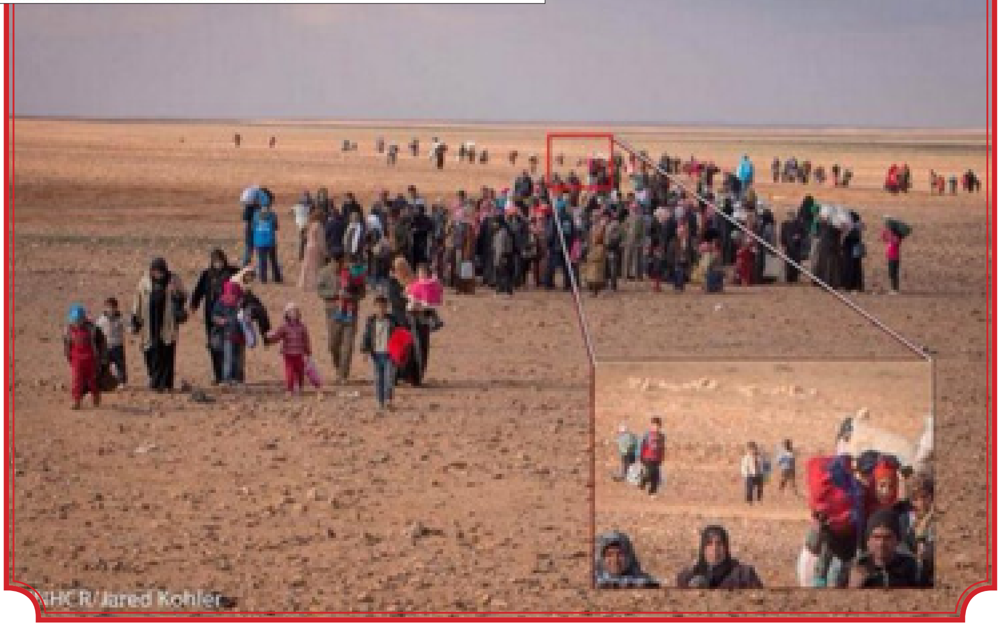
[3] www.bustle.com/articles/15809-syrian-boy-found-alone-in-desert-by-un-after-becoming-separated-from-fleeing-parents-photo

[4] www.cjr.org/behind_the_news/syria_not_orphan_boy_pic.php

[5] www.theguardian.com/world/2014/feb/18/image-syrian-boy-desert-un-refugees-tweet



Andrew Harper @And_Harper · 18 Feb 2014
Thanks to Jared 4 this shot showing **Marwan** at the back of this group of @refugees. He is separated - he is not alone.



And two days after that original tweet, Andrew Harper clarified; Marwan was at the back of a larger group of refugees, not, as many understood, crossing the desert 'alone'.

Using 140 characters to describe a complex situation

In the time between Harper's first and second tweet, thousands of people saw that photo, and interpreted the message to mean that 'temporarily alone' meant more than just a few feet behind his family. It was picked up in **Time magazine**,[6] where, ironically, the photo was described as "*an image cut[ting] through the fog to illustrate a simple truth in a way no amount of words or numbers ever could.*".

Using social media to garner support for humanitarian activities is understandable; but this example held the risk of misrepresenting what was, it seems, a simple case of **a boy standing 30 feet behind his family**.[7] Though having scepticism about the 'validity' of anything we see on social media is healthy, this case brings up a number of responsible data issues:

[6] time.com/8359/syria-refugees-toddler/

[7] https://twitter.com/jimsciutto/status/435743490417233920?ref_src=twsrc%5Etfw

CONSENT

Were Marwan's family aware of what was being tweeted? Though his name was changed, the photo remains online in many different places. Without solid technical literacy–or first-hand experience of Twitter, and its global reach–it may well be hard to imagine just how far this photo would reach–from the thousands of retweets on Twitter, to the republication of the tweet in major media outlets.

IMPLICATIONS

"Briefly separated", with a photo of a little boy on his own in the desert, was evidently understood by many to signify a more serious separation than what the reality seems to be–that the boy was walking within sight of his family. The more serious question (and **a debate ensued on Twitter**[8] along these lines) was whether this misrepresentation was intentional, potentially aimed at using the situation to draw attention to the undoubtedly crucial role of UNHCR staff, or if it was unintentional, written quickly and with the obvious limit of just 140 characters.

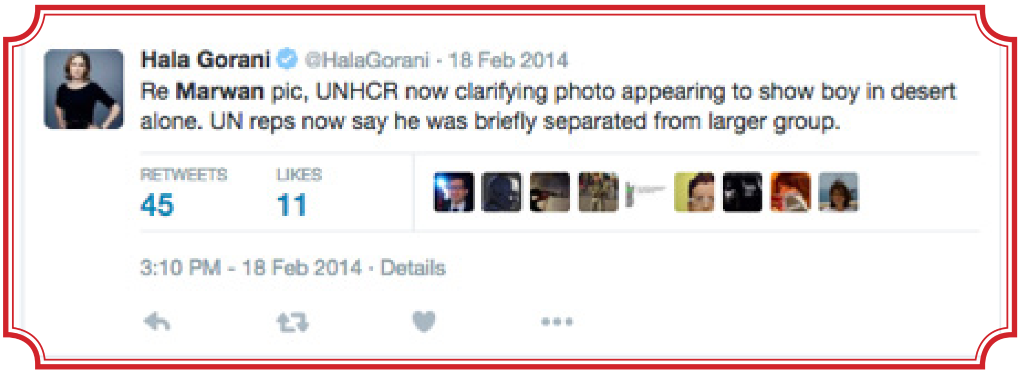
VIEWS OF INDIVIDUALS VS ORGANISATIONS

The original tweet was sent out from an individual's account, not from the official UNHCR one. Despite this, discussions online seem to assume that whatever was said was an "official" UN perspective. Would a disclaimer in Harper's bio have affected this? It seems unlikely, but would it have made a difference to the way the story was reported? Or is anything tweeted by senior members of staff considered to be an official representation of the organisation's views?

Representation

Although in this case, Marwan himself was not in as serious a situation as many understood him to be, it is true that there are many other children who are in that tragic situation. Despite the lack of veracity of this specific photo, **Sara Gates writing in the Huffington Post**[9] still describes it as "epitomising the Syrian refugee crisis", and **others found the silver lining in the situation**[10] as "throwing a spotlight on the number of child refugees". Is it ever okay to use one example to be 'representative' of a broader situation as a whole?

[8] **https://twitter.com/jimsciutto/status/435743490417233920?ref_src=twsrc%5Etfw**
[9] **www.huffingtonpost.com/2014/02/18/syrian-boy-desert-marwan-separated-family_n_4808576.html**
[10] **www.theguardian.com/world/2014/feb/18/image-syrian-boy-desert-un-refugees-tweet**



Hala Gorani @HalaGorani · 18 Feb 2014

Re **Marwan** pic, UNHCR now clarifying photo appearing to show boy in desert alone. UN reps now say he was briefly separated from larger group.

RETWEETS 45 LIKES 11

3:10 PM - 18 Feb 2014 · Details

Mitigation strategy

A part-retraction was issued by Andrew Harper, the person who tweeted the original photo, thanking his colleague Jared Kohler for a photo which put Marwan's positioning in context with the rest of the group. CNN anchor Hala Gorani, whose retweet gathered more attention than the original, also tweeted a slight clarification:

But again–'briefly separated' can be interpreted in a wide range of ways. She deleted the original tweet, but it's also worth noting that this clarification got only 45 retweets in comparison to the original which received up to 10,000 retweets before it got deleted.

Updates were also posted on the TIME piece,[11] though many others did not adjust their original post. Mainstream media picked up upon the misunderstanding too, with the Guardian saying **it first triggered sympathy, then a backlash**,[12] and **the Independent in the UK covering the explanation**[13] that Harper subsequently tweeted.

Lessons learned

As **Sara Morrison writing in the Columbia Journalism Review writes**,[14]

> *"The news is supposed to give its consumers information, not create fabricated narratives... Lie in one photograph (even a lie by omission, as in Marwan's case) and cast doubt on all of them."*

[11] time.com/8359/syria-refugees-toddler/

[12] www.theguardian.com/world/2014/feb/18/image-syrian-boy-desert-un-refugees-tweet

[13] www.independent.co.uk/news/world/middle-east/syria-crisis-image-of-four-year-old-boy-marwan-crossing-into-jordan-captures-plight-of-refugees-9136290.html

[14] www.cjr.org/behind_the_news/syria_not_orphan_boy_pic.php

For media reporting on these tweets, fact-checking them before quoting seems to be the main lesson learned from this situation. Though Harper **seems to stand by his comments**,[15] this case acts as an example of what can happen if potentially ambiguous situations are covered without total transparency.

Clarity, despite the inherent lack of space in Twitter, seems crucial. Corrections from media agencies upon realising that the story might not have been covered correctly, appear also to be one of the more responsible ways of dealing with faulty coverage.

Read more

'**The photo that cried wolf**', by Sara Morrison in the Columbia Journalism Review[16]

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room's Responsible Data Program, 2016.

[15] **https://twitter.com/And_Harper/status/435652401048350720**

[16] **http://www.cjr.org/behind_the_news/syria_not_orphan_boy_pic.php**

Responsible Data Reflection Stories 6

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Building a tech tool for sensitive data

Collecting data on Sexual Violence in Conflict Zones

Context

In 2011, US-based organisation Physicians for Human Rights[1] launched the Program on Sexual Violence in Conflict Zones, after identifying that many survivors do not come forward to report their cases. They began by working with a variety of stakeholders who are each responsible for some aspect of documentation around the incident, to improve their technical skills in documenting and preserving forensic evidence to support local prosecutions of the crimes.

Their responsible data challenges centred around digitising the process of communication and information collection between these partners, bearing in mind that the data they are collecting and working with is of a very sensitive nature. They engaged in a thoughtful process of assessing needs and working out what needed to happen before engaging in any technology development, then iterating upon the application, Medicapt, that was developed.

[1] http://physiciansforhumanrights.org/

Based on an interview with Karen Naimer, Director of their Program on Sexual Violence in Conflict Zones, the challenges they faced and mitigation strategies they engaged with are related below.

BACKGROUND

In conflict zones, few survivors of sexual violence end up coming forward to report their experiences, for a number of reasons; stigma, fear of reprisals, or not knowing who to turn to, for example. The Program on Sexual Violence in Conflict Zones is aimed at helping those survivors who do come forward increase the likelihood of successful prosecutions by training clinicians and other first responders to more effectively collect, document, and preserve forensic evidence of sexual violence to support allegations of these crimes. The programme focuses its efforts in the Democratic Republic of Congo, and Kenya.

The programme began by bringing together people from various sectors who are involved in the process of gathering evidence to support a case: doctors, nurses, police officers, lawyers and judges, so that they can explain to each other their specific roles and responsibilities. Through multiple in-person workshops, they realised that documenting evidence of sexual violence to be used in court needed multiple steps from each of the stakeholders.

They realised that there were many obstacles to be overcome in order to better coordinate this collaboration, which at the time, took place via sharing information on paper between multiple stakeholders; transport is limited, the roads are terrible, and the cost of paper, photocopying and printing was very expensive.

The challenges

STANDARDISING DATA COLLECTION

Given that mobile phone penetration is high in the regions in which they are working, especially Kenya, they began to think about a way of using mobile phones to share the necessary information. Their first step was thinking about how to standardise the information collected as at the time in the DRC, there is no nationally adopted standard document or medical intake form for sexual violence. Before moving further on the 'technology' aspect of the project, they identified this gap and through a series of in-person workshops and iterations, they worked closely with local partners to understand what the different needs were, and in what format data should be collected. They then worked with their cross-sectoral network to develop a forensic medical certificate that would focus on very specific needs that lawyers and judges would need to see for these cases.

This meant that the information required as medical evidence could be reduced to a single document that the doctors could submit as evidence in court, and that police could use in their investigations–essentially, making it easier for relevant data to be collected in a useful format, by all parties involved in the process. Having the hard copy printed form was the first stage of this process towards collaboration on data collection, and it is now being used in hospitals in North and South Kivu.

DIGITISING DATA COLLECTION

As a first step, they looked at off-the-shelf technology options that might have been suitable for their needs. Working with a company responsible for a survey platform used in public health situations, they put their certificate into the already-developed platform, and tested it out in January 2014.

It turned out that putting their standardised medical certificate into the platform resulted in a very cumbersome process for clinicians who had to use it. They field tested the platform with 7 clinicians from different hospitals in Eastern DRC, who were already on board and committed to the project. From this experience, they learned two major things: that convincing them of the benefits of the technology was not an issue, but that this platform in the format it was in, was not feasible.

ITERATING UPON THEIR TOOL CHOICE

To work out what would need to be in a useable and effective digital platform, and what needed to change from what they currently had, in January 2014 they held a 3-day workshop with people from their network, asking them: what would be in your wishlist for this technology, and what would the 'must-haves, should-haves, and could-haves' be, for the technology to be useful? Participants suggested ideas that they had expected, but also many that they had not anticipated. They collected all of these ideas, which then informed the process for the next phase of development.

From January 2014 onward, they undertook a big landscape analysis of existing technology tools at that point, to see what off the shelf options there were to identify a new model for this platform. This included interviewing many different technologists, privacy experts, and those working in public health, and as a result, they realised that no existing technology options met the specific needs that they had identified, and so that they needed to develop their own technology.

By the summer of 2014, they had identified a software development company to work with, and sat down with them to design and develop a mobile platform that met the needs of their end-users in the field. Their initial idea was to pilot and test the platform in communities they had already been working with, but with the possibility of it being adapted for global use.

REALITY CHECK

A year later, in January 2015, they went back to the same clinicians who had taken part in the field test a year previously, as well as a couple of new people, and introduced them to Medicapt 2.0. The clinicians felt ownership over the design and the application when recognising that their comments had been prioritised, and that all of the features that had been classed as "must-haves" were present in the new version.

Lessons learned

The process of developing Medicapt is still ongoing, but at the time of talking, Karen had already identified a number of key lessons dealing with making decisions around technology.

Looking around for existing tech solutions was helpful for them to work out exactly what they wanted from the end product, even though they didn't end up using one in the end.

Deciding to redirect their course wouldn't have been able to happen without the support of funders who trusted their decisions, and they were able to back up this decision thanks to ongoing feedback from end-users. Making that decision took courage, too–being able to flexibly change plans, budgets, team allocations and their strategy, wasn't easy, but they realised that having the priority in mind of building something actually useful was far more important than sticking to their original plan.

Finding a development team who really understood the sensitive environment in which they were working, was also tricky. They took their time in choosing a company, and spoke to a lot of people, and in the end, were very happy with the company they ended up partnering with: Karen recommended taking time to find the right home for the product, rather than rushing into any partnerships, and being very clear in explaining their needs. They also made sure to have ongoing, almost constant communication with the development team–even bringing one of the developers to the region to field-test the product.

Throughout the process, they have been working with the mantra of making mistakes early on in the process, and redirecting based on these mistakes. They also tried to bring end users of the product into the process from the earliest stages, rather than waiting for any kind of "perfect" product to share.

Given the sensitive nature of the data shared and collected, they have been in conversation with a number of technologists and security experts, and even though they've had a number of ideas of features that could make the product better, they seem to be waiting for security solutions before implementing them.

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room's Responsible Data Program, 2016.

Recognising uncertainty in statistics

Quantitative data is the result of numerous subjective human decisions.

Context

> *More often than not, it is not the writer that is twisting the numbers but the numbers themselves twisting up the writer; manipulation of the facts, or of the reader, is usually not intentional. The exploration of the use and misuse of numbers is at the base of a large, and growing, body of academic and popular work on quantitative literacy.–p3, Numbers are Only Human, Brian Root*

Understanding and using statistics responsibly in human rights advocacy can be incredibly difficult. As Brian Root excellently outlines in his article, "*Numbers are Only Human*"[1], understanding that quantitative data is the result of numerous subjective human decisions, can make a big difference to how an organisation chooses to use certain statistics to support their work.

One concrete example of how difficult these decisions can be, though, can be seen if we look at data on killings due to the ongoing Syrian conflict.

[1] **https://global.oup.com/academic/product/the-transformation-of-human-rights-fact-finding-9780190239497?cc=us&lang=en&**

Background

There are several organisations who are monitoring casualties in Syria–including, but not limited to[2]:

- March 15 Group
- Syrian government
- **Syrian Center for Statistics and Research**[3]
- **Syrian Network for Human Rights**[4]
- **Syrian Observatory for Human Rights**[5]
- Syrian Revolution General Council
- Syria Shuhada Website
- Violations Documentation Centre, the documentation arm of the Local Coordination Committees
- Damascus Center for Human Rights Studies

Figures released by all of the above parties (some of whom are no longer collecting up to date data) differ greatly. As an example, in 2013 the Syrian Observatory for Human Rights (SOHR) enumerated estimated that the **TOTAL NUMBER OF PEOPLE KILLED WAS 110,371 PEOPLE**.

[2] Selection taken from HRDAG report, **https://hrdag.org/wp-content/uploads/2013/06/HRDAG-Updated-SY-report.pdf** with additions

[3] **http://csr-sy.org/**

[4] **http://sn4hr.org/blog/category/victims/death-toll/**

[5] **http://www.syriahr.com/en/**

Estimate vs Enumerate

Using a convention from the Human Rights Data Analysis Group (HRDAG), the word "estimate" is used to describe an estimation of a figure that includes a measure of uncertainty. Other non-statistical 'guesses' are referred to throughout this piece as "enumerations".

They even released more disaggregated figures along with this estimation:

Civilians killed	40146 of which:
women	nearly 4000
children	more than 5800
Rebel fighters	21,850
Regime army soldiers	27654
Pro-regime militia	17824
Hezbollah	171
Unidentified	2726

THE PROBLEM OF SUBJECTIVITY

As Root identifies, the disaggregated categories within the data are also subject to a lot of human decision making:

> *Imagine the decision that might have to be made to categorize a typical citizen with no military training, who has picked up a gun shortly before his death. Perhaps the coder might have a bias to continue calling this person a civilian. But this person took up arms against the government, did they not? How would you code a Syrian army defector now fighting with an opposition group? ... Without some sort of standard protocol, rigorously followed, the coding of affiliation allows for a degree of subjectivity...–p6*

As he rightly identifies–there are lots of human decisions that go into creating these statistics, and without knowing how these deaths have been coded, it's difficult to trust in the figures. But this nuance can be difficult to convey without using long-winded explanations, and sadly, soundbites of short, snappy figures, often get much more traction in public debate (see **Reflection Story 6** for more on this).

The messiness of the Syrian conflict adds to this confusion in coding; it can easily be unclear who is responsible for a certain attack, especially those which nobody wants to take responsibility for, such as chemical weapons attacks.

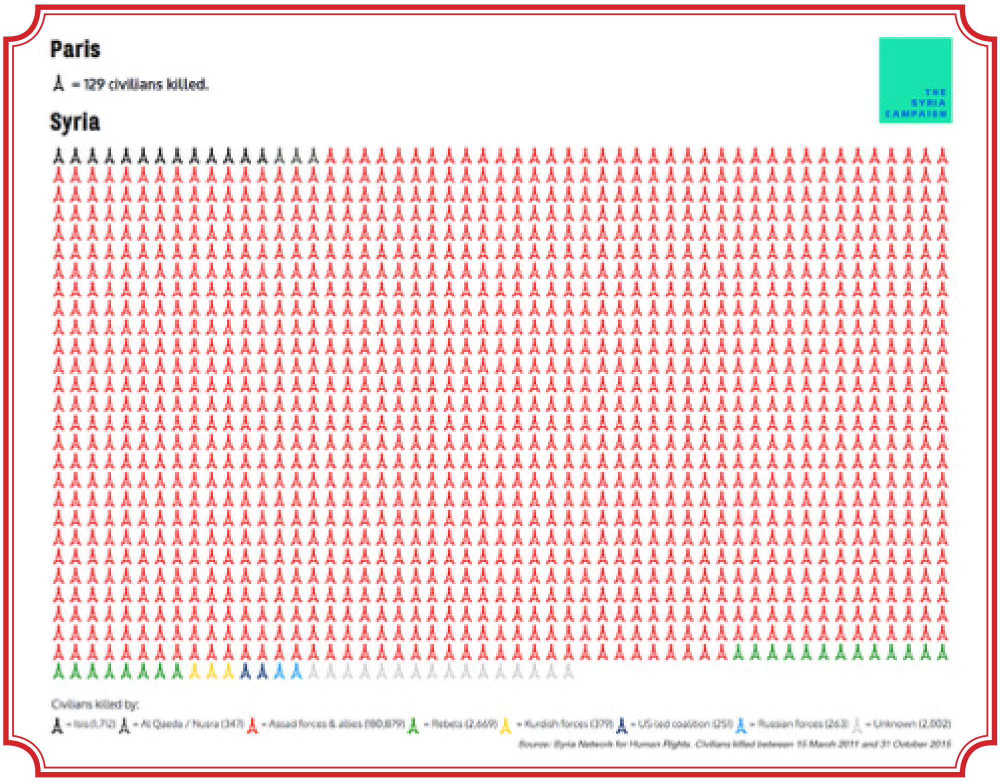


Image from the Syria Campaign: **https://thesyriacampaign.org/**

A FALSE SENSE OF ACCURACY

As Anita Gohdes, researcher at the Human Rights Data Analysis Group, highlights: in addition to the problem of subjectivity and human judgement, this level of disaggregation can convey a false sense of accuracy. Breaking down figures to such precise levels–"21,850 rebel fighters" rather than "Approximately 22,000 rebel fighters" removes the sense of uncertainty that is undoubtedly there.

It's true though, that images such as the visualisation above draw attention to some important issues. Though they state their data source (the Syria Network for Human Rights) what we've explored here so far makes it clear that **THIS DATA HAS FLAWS**. We can't know for sure the extent of those flaws, though, and some might argue that as long as the main message is transmitted, the details don't matter so much.

A data-focused group, the Human Rights Data Analysis Group, (HRDAG), published in mid 2013 a statistical analysis of documentation of killings in Syria[6], commissioned by the Office of the UN High Commissioner for Human Rights. In it, they integrate findings from eight different databases; seven built by Syrian human rights monitors, and one from the Syrian government.

[6] **https://hrdag.org/wp-content/uploads/2013/06/HRDAG-Updated-SY-report.pdf**

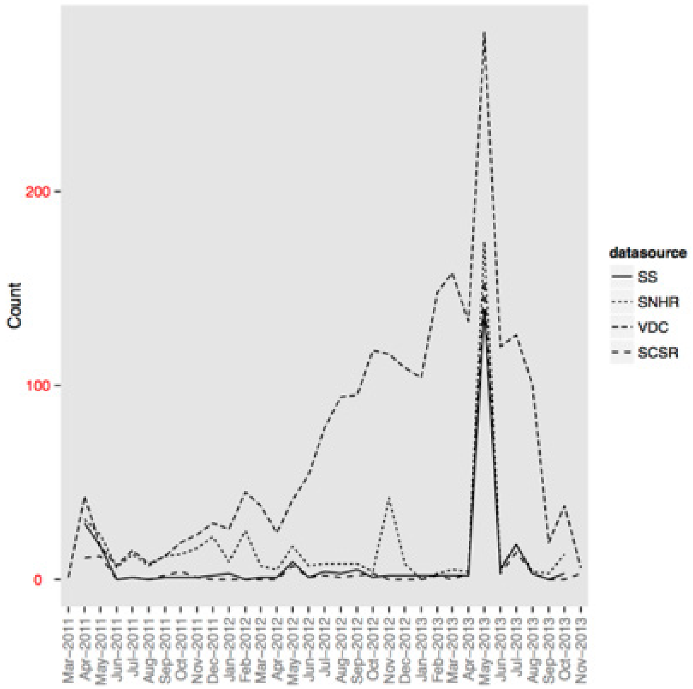


Image taken from **'Searching for Trends: Analyzing Patterns in Conflict Violence Data'** post by Megan Price and Anita Gohdes, copied here with permission

In it, they qualify carefully their findings; that the final figure they came up with, 92,901 unique killings of both combatants and non-combatants, is an enumeration and not the complete number of conflict-related killings. They identify potential problems like undetected duplicate deaths among the different databases; inaccurate records within any of those databases; victims presumed dead who may have later been found alive; or, undocumented killings, those that don't appear in any of the the eight databases for any number of reasons, to name just a few.

To highlight the problems that they discuss, take a look at the graph above. It shows **THE DAILY COUNT OF UNIQUELY REPORTED KILLINGS** in the area of Tartus, as collected by four well-known data sources.

As you can see, there are some big discrepancies in the data, notably leading up to May 2013.

> *All four sources depict a marked increase in violence in* May 2013[7]*, which corresponds to an alleged massacre in that governorate. Three of the remaining sources observed relatively few victims outside this single spike in violence. The fourth source, VDC, describes the observed peak in May 2013 as the culmination of steadily increasing reports of violence throughout the preceding year. If we did not have access to the VDC data, we would erroneously conclude that there is consensus among data sources that relatively little violence is occurring in Tartus, and that May 2013 was a relatively isolated event.* –'Searching for Trends: Analyzing Patterns in Conflict Violence Data'[8] *post by Megan Price and Anita Gohdes, April 2, 2014.*

The challenge: admitting weaknesses in the data, while pushing a strong message

As HRDAG and Brian Root at Human Rights Watch have identified, getting "accurate" data on killings in the Syrian conflict is incredibly challenging. Any and all of the major data sources face uncertainties in getting the data, in coding it accurately according to type of death or killing, and in comprehensive data coverage of hard to reach areas. In short; none of the data is certain accurate enough to be cited as "truth" or fact.

It's true, though, that data on the topic is very much needed for a number of reasons; in order to get an idea of the scale of the conflict, to understand what humanitarian needs there are; and on the advocacy side, to get people's attention to a tragic situation, and garner public support.

There are many ways of counting deaths in conflicts that have already ended, as outlined **in this Guardian article**[9]–and even when they have ended, there is still a great deal of uncertainty around the accuracy and subjectivity of that data. This challenge is even greater in the case of Syria, where the conflict is ongoing and increasingly messy between different stakeholders.

The main challenge in this case is knowing where to draw the line between using quantitative data to strengthen advocacy, or where to admit uncertainty and potentially weaken the key advocacy message.

[7] www.bbc.com/news/world-middle-east-22410392

[8] politicalviolenceataglance.org/2014/04/02/searching-for-trends-analyzing-patterns-in-conflict-violence-data/

[9] www.theguardian.com/global-development-professionals-network/2015/sep/08/from-syria-to-sudan-how-do-you-count-the-dead

Ways of dealing with uncertainty

The Human Rights Data Analysis Group are pioneering the way in collecting and analysing figures of killings in conflict in a responsible way, using what's called '**multiple systems estimation**'.[10] But it's true that the statistical skills required to use these kinds of techniques may well lie far beyond the reach of many advocacy groups, and in this case, there are a few other considerations and techniques that can be employed:

- questioning methodologies of how the data has been gathered and analysed– for example, who decides what code is given to a data point? What levels of verification do they have in place?
- making proactive decisions around whether to include or focus upon a certain statistic as part of the advocacy, or to include qualifiers about how reliable it may or not be (eg. adding 'estimated as of ____')
- linking to other sources of deeper enquiry of the reliability or consistency of the data (eg. in this case, the HRDAG study linked above)

Clearly, with an issue as complex yet important is this–and, as with many of the reflection stories- there is no **RIGHT** answer. At the heart of mitigating against misunderstandings of data is increasing the level of data literacy of those who will be looking at the visualisations, but this is easier said than done.

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room's Responsible Data Program, 2016.

[10] **https://hrdag.org/wp-content/uploads/2013/04/Manrique_Price_Gohdes_WorkingPaper.pdf**

Responsible Data Reflection Stories 8

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Data generated by mass campaigns

The case of a public consultation around "over-the-top services" and net neutrality in India.

Context

In April 2015, the Telecom Regulatory Authority of India (TRAI) launched a public consultation around "over-the-top services" and net neutrality in India. In response, a group of net neutrality activists started a campaign–**Save the Internet**,[1] to mobilise the public to respond to this consultation, and advocate for net neutrality in India.

The campaign was far-reaching, and managed to engage large groups of people in caring about net neutrality, through light-hearted posts like "**Why does #SavetheInternet hate free?**",[2] and using innovative campaigning tactics like **gathering memes**[3] for others to use, and **funny videos explaining net neutrality**.[4]

[1] http://www.savetheinternet.in/

[2] http://blog.savetheinternet.in/why-does-savetheinternet-hate-free/

[3] https://goo.gl/YSJkdr

[4] https://www.youtube.com/watch?v=mfY1NKrzqi0

As a result, the campaign resulted in over **1 million emails sent in response to the public consultation in just 12 days**[5]- far surpassing the number of emails ever received by the TRAI previously. They reached their advocacy aim, mobilised a massive population to speak out against suggested proposals that would have violated net neutrality, and made their opinions strongly known to the TRAI, the group that would be deciding upon the future of the internet in India.

The challenge

But unfortunately the campaign took an unexpected turn. After receiving over 1 million emails, the TRAI took the decision to **PUBLISH ALL NAMES AND EMAIL ADDRESSES** of the people who had written to them as part of the campaign, on their website, organised by date sent.

In one single PDF file, they published in cleartext, personal details of all of the engaged citizens who had chosen to take part in the public consultation. They also published the content of the emails they had received–so people emailing them with personal information in their email signature, also had that published online.

As a result, **over 1 million email addresses, together with associated names, and the dates they emailed the TRAI**,[6] were made available on their site. Essentially, the TRAI created a treasure trove for spammers or those interested for whatever reason in harvesting large datasets of people's names and details.

In response to this, a group identifying as an Indian spinoff of Anonymous, AnonOpsIndia, **carried out a DDOS attack against the TRAI site**,[7] managing to bring it down for two days, though the **TRAI denied any such hack, and said the site was down due to 'technical glitches'**.[8]

Lessons learned

The volunteer-based group coordinating the Save the Internet campaign could not have known that the TRAI–a government entity charged with protecting consumer interests–would have taken the bizarre, and dangerous, move of publishing the emails they received. The public responded to the decision online, and **TRAI faced huge criticism for publishing the information**.[9]

[5] http://tech.firstpost.com/news-analysis/net-neutrality-deadline-trai-receives-over-million-emails-from-netizens-asking-to-save-the-internet-264548.html
[6] http://www.huffingtonpost.in/2015/04/27/trai-publishes-emails_n_7149658.html
[7] http://www.hindustantimes.com/tech/trai-reveals-a-million-net-neutrality-email-ids-gets-hacked/story-0baEwoudT9uOx2sNI8fF5N.html
[8] https://thehackernews.com/2015/04/net-neutrality-trai-emails.html
[9] http://timesofindia.indiatimes.com/tech/tech-news/Net-neutrality-Trai-exposes-1-million-email-IDs-to-spammers/articleshow/47067807.cms

In response to the backlash they received for publishing the data, the **Indian Express then reported the TRAI as saying**:[10]

> *"all stakeholders are hereby informed that during submission of their counter comments, if anyone desires that his/her email id should not be displayed, it may be specifically stated so in the email... Such respondent should also include 'Do not display my id' in the subject of the email, "*

Their attitude towards publishing personal information, then, seemed to be that of an 'opt-out' of publishing, rather than opt-in. In the absence of more technically literate policies and responses from government agencies, it's difficult to envision what could have been done differently in this case.

For future campaigns, Save the Internet advise people to do as the TRAI says, and 'opt-out' in their emails responding to public consultations–but a consultation has not yet been completed since then, so there's no proof that they will indeed take note of those who choose to opt-out.

the engine room

This publication is part series found at **https://responsibledata.io**, produced by the engine room's Responsible Data Program, 2016.

[10] **http://indianexpress.com/article/technology/technology-others/mention-do-not-display-my-id-if-you-want-your-net-neutrality-emails-to-remain-private-trai/**

RESPONSIBLE DATA REFLECTION STORIES 9

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Opening the wrong data

The case of an open data hackathon and the police force.

CONTEXT

In early 2015, a group of open data advocates organised a hackathon, in partnership with a number of local government bodies, and a state police force. Hackathon organisers encouraged their state partners to "open" their data–that is, to publish it online in a machine-readable format, under an open license.

Hackathons partnering with government entities is not uncommon–for the government bodies, they provide good opportunities to see their data being used by their 'target audience', and for developers. For those attending the hackathon, the opportunity to access previously unavailable data is an added incentive to attend, as it potentially opens up the opportunity to gain new insights about their societies.

The challenge

A police officer from the police department made the decision to open up two of their datasets for those at the hackathon to use; one on mugging data, and another on mobile phone calls records. This second dataset contained a lot of personally-identifiable information; metadata of when calls took place, to whom and from whom, and their location.

Though the police force had been aiming at helping the hackathon organisers to get useful data for their event, the person responsible for this had inadvertently published online a dataset that should not have been opened, putting it available for download on their site. A number of attendees at the hackathon downloaded the data, not all of them realising the implications of the data that they then held.

A group of privacy advocates, who had heard that the hackathon was involving the government bodies, intervened after seeing the data that had been published. By going to the venue of the hackathon, they convinced the police force to take down the dataset, though it still took a few hours to do so.

In addition to the risky content of the data, it turned out that the data had been gathered as part of a terrorism investigation; so, privacy implications for individuals aside, potentially also held security risks for publishing.

Lessons learned

The hackathon organisers resolved to make sure that only data that should be opened should be published as a result of their open data advocacy; but things aren't quite that simple. Given that the police force heard their calls to open data, then published the data on their own site, there was effectively little that they could do after realising that inappropriate data had been made public, except for asking them to take it down.

Clearly, in this case, the person/persons charged with choosing data to open from the police force were unaware of the consequences of their actions, and there is little that a group of hackathon organisers can do about this except for offering more advice where needed.

A better lesson, perhaps, would be to be very clear in initial contact with government entities as to what kinds of datasets they might be looking for, and asking for a first opportunity to 'check' the data prior to publishing. The risk then lies in creating too many barriers for them to overcome before publishing, but on balance, especially with institutions dealing with sensitive data, perhaps this is a preferable outcome.

The police force in question never mentioned the mishap publicly, instead choosing to take the data down from their site without comment. The fact remains, though, that a number of people had downloaded it during the few hours it was online, and if it hadn't been for the watchful eye of privacy advocates in the region, the data may well have stayed online for much longer, with unforeseen consequences.

the engine room

This publication
is part series found at
https://responsibledata.io,
produced by the engine room's
Responsible Data Program, 2016.