

Transcript: Q&A

Using Big Data to Understand Global Conflict

Kalev Hannes Leetaru

Senior Fellow, Center for Cyber & Homeland Security, George Washington University

Chair: Caroline Baylon

Research Associate in Science, Technology and Cyber Security, International Security Department,
Chatham House

24 April 2015

The views expressed in this document are the sole responsibility of the speaker(s) and participants do not necessarily reflect the view of Chatham House, its staff, associates or Council. Chatham House is independent and owes no allegiance to any government or to any political body. It does not take institutional positions on policy issues. This document is issued on the understanding that if any extract is used, the author(s)/ speaker(s) and Chatham House should be credited, preferably with the date of the publication or details of the event. Where this document refers to or reports statements made by speakers at an event every effort has been made to provide a fair representation of their views and opinions. The published text of speeches and presentations may differ from delivery.

Caroline Baylon

I actually have a question for you. You mentioned ranking the world's heads of state by popularity. Who came out on top?

Kalev Hannes Leetaru

Good question. I'd have to look. It actually updates every day, although we're going to come out shortly with a new version of it that is even more powerful. It was actually very interesting because oftentimes what you saw was probably 90 per cent of world leaders were in the red, they were negative coverage. If you think about it, usually the news media doesn't – there's not usually an article that says our head of state is doing a great job, we should all applaud him or her. Usually news coverage is they did something that people disagree with. It's just basically, are people really upset with that person or are they just kind of upset with that person? The only people who make it into green tend to be, for example, where a country got a big aid package or a big bailout, or there's a new president and everyone is really excited for a few days until the person actually has to make their first decision.

So it's very fascinating to see. What's interesting to me is the changes day to day, where you see this person is becoming more and more negative. Or this person, in the case of Assad, for example, what you're actually seeing, if you start really drilling into that map, is basically anytime there's a huge slide of the media really darkening about him, it's when he does big attacks. Whether the Ghouta chemical weapons attack, whether some of his barrel bombings and other things. Of course, he's not watching the world's media and saying, wow, everyone is being bad about me, I should do it. It's more about the media is capturing a dimension that is really powerful.

Caroline Baylon

Thank you. We're going to take questions now from the audience.

Question 1

It's really magic, but unfortunately my nation never talks about how to do magic. Until last summer, we could use [indiscernible] wiki – it's a free software, just 15 pounds. So we can use as journalists. But Twitter shut down. So do you have any alternative on how to scrape Twitter and blogs and Facebook? If there is some product we can buy in the market, please tell us.

Kalev Hannes Leetaru

GDELT actually, all the data I produce is openly available. The articles are not, but basically all the information. So for example, the list of all the riots that it's identifying in the news each day. The list of what it's finding around the world, that is openly available as .csv files, anyone can download it and map it and do everything.

I've done a lot of work in social media. Right now GDELT does not incorporate social live, mostly because Twitter and Facebook are obviously big in many countries, but Burma and other places, you have for example [indiscernible], you have Weibo, you have all these different platforms. We start moving beyond purely the west, many of these platforms are more difficult to get content from. Twitter, I've got some forthcoming work that shows the degree to which Twitter is being Astroturfed today. There are nation-states and other actors – I'm not talking about ISIL/ISIS/Daesh, they're doing everything overtly. They're

publicly putting stuff out. There are a number of nation-state and other actors that are manipulating Twitter to a very strong degree. In the old days it was simple, you looked for a Twitter account that said 'XYZ 700' and it's only sent 10 tweets and they're all about one thing – like for example, in the US both the upcoming election and the last election, the presidential candidates have all got in trouble for buying Twitter followers and buying accounts that would tweet good stuff on their behalf. Today these things are becoming incredibly sophisticated.

So to detect those clusters requires significant advances over what we have today to be able to detect these clusters, because these clusters exist. They're very diffuse, they're very difficult to detect unless you're looking at all Twitter over periods of time, doing a lot of expensive calculations. So right now, I do not incorporate those. But this is, of course, an issue with Twitter, Facebook and others. Increasingly, social media is only available if you have a paycheque. Of course, that is a big problem. We oftentimes think about social media in particular as a town square, it's ours. But as we're finding, Turkey and a number of nations are – for example, one particular nation was getting very good at DCMA takedown requests. So basically saying any anti-government post that's up, the way in the US – and most of these companies are headquartered in the US – all you have to do is a DCMA takedown request. Under US law, if that company removes that post, it doesn't matter. If they leave it up and it turns out even if it's just someone saying 'I think the president's an idiot' and they say they're living in Turkey, if a US company removes that post, even though it's pretty clear there probably isn't a copyright call on that, basically US law is kind of pathologically rewarding countries to increasingly interact with the social sphere in a very interesting way.

I know that doesn't quite answer your question. That is a difficulty. These things are increasingly becoming more difficult to access both for research but also – VK, a lot of the anti-Putin protestors, there's no longer – the VK presence of the anti-Putin protestors is becoming more diffuse. You're seeing more movement, constant fluid movement. So I think social presents some very interesting challenges in trying to understand the world.

Question 2

I did use some software in late 2001 to predict a major fall in sales tax revenues in Britain, in the hotel sector, because there would be a significant decline in international travel in late 2002 or early 2003, because that's when America was going to invade Iraq. But I would point out that the software I used was this. And the fact that the American president in 1991 was called George and the American president in 2001 was called George. A quick question but firstly just to say, also, I'd like to offer my condolences and commiserations on the World Bank. I worked on a World Bank project in Yemen and it was rather difficult getting the international World Bank staff, not the Yemeni World Bank staff, to take into account that the work needed to be modified to take account of the fact that the country was moving towards civil war. A different situation.

The question is: what work is being done – you did hint at this as a need – to marry up the value of what the quantitative collection and the quantitative listing and analysis can do, with existing think tank, Chatham House type of predictive analysis? Are the various groups being brought together? How are they getting on with each other?

Kalev Hannes Leetaru

It's a great question. The data – it's interesting, if you think about it, there are many patterns that are largely self-evident. So for example, there's a terror attack somewhere – you might anticipate tourism in

that area – for example, Egypt is having increasing instability. You might say there's probably going to be potentially a decline in western tourists going to the areas that are most unstable.

But what becomes interesting is – people talk about black swans. They say so much of what's happening around the world is completely unpredictable. We could never have seen this, because it's so rare. Especially country collapse. A lot of people say country collapse is entirely unpredictable, you could never predict this, mostly because you don't have enough training data to train your models on. What I say to that is: yes, whole country collapse is very rare, but countries come to the brink every day of that. They come very close to that brink or certainly they become far more unstable. Look, for example, at the Arab Spring. There were a number of countries that should have potentially gone under, based on all the other characteristics that were there. Some of those, other neighbouring nations intervened to halt the slide. But every day, countries undergo various levels of instability. Yes, probably right now we're not at the point where we can actually predict entire country collapse, although some of my previous work actually did show – if you look at the media tones, if you look at worldwide media coverage about a country and you look at the tone of that, if you basically look for points where you essentially drop off a cliff, the country basically goes under within about one to two weeks. If it doesn't, if it's stable, the country is largely stable. This actually holds in Libya. You get the restabilization and the stalemate.

So there are a lot of indicators. They're short-term indicators. But there are a lot of indicators that can happen from this. I can say there is some forthcoming work that is coming out on this that shows the overprint of conflict onto daily societal life. So people make a lot of subconscious decisions about actions that they take based on their understanding. People don't act based on facts, they act based on their interpretation of what they think is going to happen around the world. The simplest case is people are hearing the news that Yemen is falling apart, so they cancel their vacation to Yemen. That's a simple case. In Yemen's case, that may or may not be a bad idea.

There's a lot of subtle things that we have not yet displayed. Some of those graphs that I showed, there's a lot of hidden patterns in data that we've not seen. So especially if we're trying to do forecasting – for example, if you work with firms on Wall Street, you see the price of wheat down in this little corner of Zimbabwe gives me the price in London [of gold] at four days' notice. There's so many of these patterns that are being found on Wall Street right now. The price of wheat down here is not directly impacting gold. What's happening is there is this incredibly complex linkage between those.

Of course, a lot of the things that have happened here in Europe in terms of protests and other instability, a lot of these come – for example, let's say you want to look at an urban protest. Oftentimes, if you had census data that was household level in that area, that was measuring youth unemployment, theory tells you to look for things like high youth unemployment. A lot of youth at the right age and they don't perceive that they have any prospects – that's when you start seeing a lot of problems.

The problem is that oftentimes census data, for example, is measured at a large level. So for example, one area I was looking at, one country, the census block where they recorded youth unemployment included both a very poor area and a very wealthy area. In fact, that wealthy area is becoming richer. So according to your census numbers, this urban immigrant area was basically – you're seeing no change in that area, because the wealthy enclave was looped into there. If you had resolution that was household-level in real time, that's being updated more than a few times or once a year, theory would tell you to look for that.

The problem is that what you have essentially is a census tract that is essentially useless, and so you have to get the thing you want to measure – if you could actually measure that, you're perfect. The problem today is that the things that theory tells us to look for and says this is the thing you look for to tell us if

something is going to be unstable, those are things that are not measurable. Or even if you could collect the right data for that, it's so messy, it's so difficult to understand. So what you have to do is you have to say basically the thing you want to measure right here, I can't. So it's kind of like lava coming up to the surface. I know it's right here, it should come up here – it doesn't, because of something in the way. Instead it bubbles up here, here and here. Theory doesn't tell me to look here because this isn't the thing I'm trying to measure.

So this is where data becomes very powerful, because we're able to basically say: here are past times when this magma chamber has erupted, and usually it's come up here but it's also come up here, here and here. So that ability to allow data to show us the things that we're not expecting to look for because we can't measure the things we want. I think that's the greatest challenge today. A lot of the theories about conflict have not sustained well in the last few years of what's happened around the world, largely because those theories are dependent on understandings like GDP or infant mortality. A lot of theories say that infant mortality has to be skyrocketing completely for a country to fall. Obviously, we've had a number of countries that have had issues where infant mortality was certainly not on a vertical line upwards. That's because countries that fell in the 1950s, fell for different reasons than countries that are falling today, but also because it's not infant mortality that causes – it's not that people look at infant mortality ratings and say, wow, we should overthrow our government. It's because infant mortality is actually indicative of something deeper. It just happens to be the thing that we've been able to measure, the most accessible, but it's a composite of a bunch of things that nowadays other indicators.

If you think about all the things we measure, there's no such thing as a pure indicator society. Everything that we can assess, certainly in the open world, is a composite and overlay of so many other things going into it. So all these other things can shift over time.

So the short answer to that is that where a lot of this becomes powerful is we can start looking over enormous amounts of data over very long time periods, and machines are much better oftentimes at sifting out and smoothing out the noise. Also, I think the bigger part is working with, for example, NGOs that have footprints across the entire world. They've got physically people in each country. Even they oftentimes are at a loss to see what's happening because, for example, their little outpost in this corner of the country – they're seeing everything happening in this local area but they don't have the infrastructure to then send that back and send that on up the chain so that you have one master view. In one particular case, without mentioning the country organization, you had a case where this area is very stable, this area had a huge issue, and they knew about that. But what they weren't noticing is this was marching steadily this way, into an area. This group kind of knew that, but this group knew that if this conflict hits here, it's going to be a fireball. So neither side really talked to each other because most large NGOs don't have that view. Certainly being able to surface things – even if you're in-country, if you've got a few people, you're probably spending most of your time trying to do whatever it is you're supposed to be doing there, not sitting 24 hours a day monitoring the news. So the ability to kind of watch those things.

Question 3

Fascinating account, what you gave, and I certainly have no problem with the idea that more data could help to show that there are more problems in the world than the Anglophone literature and news might normally present. But if I was to think of the world from the perspective of decision-makers – as an academic, I can only imagine what that's like – I could imagine that what you're actually showing is there are more problems in the world than people previously had contemplated or understood. So in that case, how would your data be used to predict the real crises that matter, as opposed to the ones which actually,

who knows why, just go away? The equivalent of the complaint about economic analysts who allegedly predict five out of the last three recessions.

Kalev Hannes Leetaru

It's a great question. Certainly in the western world, most of the work on forecasting is basically for governments. You think about most of the work on forecasting, it's the idea that you predict where the hotspots are going to be. You inform your government, and then your government – at least in the US, it's sort of this notion of, well, you predict all the unrest, you tell the president and then the United States will come to the rescue. Even in cases where you don't need data, you know what's coming there, the US – you look at all the crises that are happening around the world, even ones that are currently happening right now that we all know exist, you don't see – no nation can go everywhere and do everything.

So one of the things that we're starting to work is we're starting to partner with a number of NGOs to explore what would it look like to try and put this data in the hands of people on the ground itself. Obviously, like the maps I'm showing, if you're sitting there on a feature phone with the equivalent of dial-up, if that, on your phone, you're not pulling up a fancy interactive map that's rendering interactively on a quad-processor cellphone.

So the question is, how do you deliver this? Because the other thing is, again, especially forecasting work, there is always a margin of error on that. So if you send a text message to this village here and you say, this group is marching up the road and they're burning everything in your area, there's a possibility they could turn away. So if you tell these people, get out of the way, they may sacrifice a lot to do that and it may be that the violence doesn't reach that area.

So part of what we're starting to explore is – I don't have the answers to this yet – but we're taking the first steps of turning the notion of forecasting on its head. Because if you think about it, the people that are in the best position to actually make use of the insights from this are people on the ground. If you say, for example, there's instability in this particular area, telling the president of the United States – the US might do something about that but more likely not, based on the politics and where the US is at the moment. Imagine being able to inform those people on the ground that basically there are tanks rolling this way, in 48 hours, if they don't deviate, they will be on your doorstep and they're torching everything in their path. I think that presents some fundamentally exciting possibilities, especially also for just helping people be more aware of what's happening. In India, for example, there's been some incredible work that's being done on the price of agriculture. So being able to tell a farmer here's what the price is in the market right now, so you can make that decision – instead of telling the government these things, the farmer is the one who's got to make the decision whether to come into market or not.

So how do you put this data in the hands of the people? Again, we're just on the beginnings of that. There's a long road to go. But to me, that's the most exciting part. That's when this stuff – because again, I've been doing a lot of work with the policy world. The World Bank just had a report out that said like 90-plus per cent of their PDF reports have never been viewed once. They did this amazing analysis of the views on all that. So if you think about the vast majority of what comes out of think tanks and NGOs and even governments, none of it ever gets read. None of it ever gets looked at. That's because, again, people have a finite bandwidth. Even when we know a situation is there, there are limited ways for us to interact with it.

So I think that's the future. I don't know what that's going to look like but I think that's basically where we're heading.

Question 4

You said in your introduction that you carried a study on how people were using social media during conflict. I would like to know if that allowed you to extract what really mattered for people in conflict situations, what were their main concerns? Were you able to analyse it, no matter the language?

Kalev Hannes Leetaru

That particular study – social media has played a role in a number of conflicts, but the 2012 Israeli-Hamas conflict was the first that truly played out on social media. It was systematic, both sides were using social media as a true campaign element. Also, ISIL/ISIS/Daesh has really taken that in a whole new direction. But this particular one, what was so interesting about it – the particular question that we were interested in was less civilians on the ground. So I also work with a number of people that do look at it from an aid perspective. That particular piece was from the notion of the combatants themselves, what's the purpose? You think about, why use social media? If you are, for example, a government, why should you even interact with something like Twitter or Facebook? Of course, the idea is that you can reach around the media, project your view and basically activate whole new communities.

One of the things we found is Hamas actually gained more than the Israeli government did in that case, but in both cases what we found – the question that we entered was, is this primarily beneficial to governments? So basically, Israel and Hamas, are they reaching fundamentally new communities, turning them on their side, basically? Essentially, force projecting.

What we found is that, at least in the 2012 conflict, at least on Twitter, the accounts that activated were people that had already discussed in the past either pro-Israeli or pro-Hamas sentiment. So essentially what you were doing was taking this vast field of latent people – we're not talking about people that are tweeting a statement every day. We're talking about somebody who maybe every few months said something here or there. What you're doing is by portioning this out, you're basically activating these existing latent communities, causing them to surge forward, but then basically they drop out. At least in that conflict, we didn't see new communities fundamentally activated, nor did we see the people who reactivated, stay activated. They basically went dormant again.

So it's a way certainly – now again, that could have had a noticeable impact maybe on how other countries interacted with that conflict. It could have driven how news media covered that. We didn't look at that. We looked purely at, did this basically cause fundamental changes in the social infrastructure around these two combatants?

Certainly in some of the more recent stuff – so from a relief standpoint, I've done a lot of work with folks like Patrick Meier and others that work on the ground on these things. My interaction with a lot of them, they've looked a lot at things like Twitter and Facebook and what's going on social. In many areas like Vanuatu or the Pakistani earthquake, in a lot of these areas there's either not social media infrastructure there or it gets wiped out when – even Hurricane Sandy did – so did the first live geographic mapping of a natural disaster, which was Hurricane Sandy. What we found was, it was very interesting. At the 72-hour mark basically, stuff started going dead. So Twitter up the coast, at 72 hours after the storm passed through, things would go dead and start marching up the coast. That was very strange. It turns out actually, in the United States, cellphone providers are required by – I don't know if it's required, but on standard, they have 72 hours of fuel in their emergency generators at their cell towers, with the assumption that within three days you can get trucks in. It turns out Hurricane Sandy was so bad it cut off

access to the roads, so the cell towers started running out of gas and dying. So Twitter was sort of silenced in a lot of these areas, not everywhere but you definitely saw a noticeable effect.

So what I'm very interested in is that combination of how do we use social media – certainly Patrick and others are doing that today, but what I'm interested in is, down the road and certainly now, how do we blend these two worlds? How do we look at, in areas where there is high social concentration? I was on a panel at the Wilson Center last year, and one of my fellow panellists said: conflict is live-tweeted now. We see everything from all the combatants. I said: that may be true in some cases. In the West Bank, portions of people were tweeting. But there's a lot of those people that are not walking around on Twitter, or certainly communications are knocked out. The Congo region, or Haiti, for example. A lot of these areas – certainly pockets exist. One of the findings of my Twitter paper was that power and Twitter correlate at 0.89. So basically, where you see power, the most remote pocket of power in a rural area, brings with it Twitter, at an interval of power. But you think about it, in the US, walking around Washington, DC – you see little toddlers with their own cellphones. I don't even know what they do with that phone, but you see little kids with cellphones. A lot of these areas, especially rural areas in certain parts of the world, it's probably the village elder that has that phone. It's not every single person in that area walking around, and certainly not live-tweeting everything around them and Instagramming it.

So again, in the future, networks keep expanding. Someday potentially everyone will be doing everything. Social media, in particular, is very good at live-breaking – for disaster purposes, it is very useful. Someone can tweet and say: we've got two people down right here, they've got this medical condition, someone please bring us medicine. Of course, that's a problem because once people can do that, then they expect medicine to arrive, and drones maybe – maybe drones can do that in future disasters, who knows.

But the bigger part is understanding the context. You see this report, this report, this report – being able to say, okay, here is kind of the boundaries of where we see cholera coming out right now. Where is this coming from? What is this leading to? Being able to actually, actively predict that stuff. We spend a lot of work on sequence detection, on macro scale things. So for example, if you have certain types of nutrition failure, and then you have a drought and you have a flood and then this. Usually, this disease comes about. So there's certain cycles that are just – or for example, in Bangladesh you've got the floods on a regular cycle. You can set your watch by it in some areas.

So part of it, I think, is how do we look beyond a single data source. A lot of my work really focuses on how do we bring all the different datasets that exist to bear. Instead of saying, let's look at Twitter, because it's easy – I can put geotagged tweets on a map, it's easy to do – how do we look across everything, to be able to say, for example – news media, for example, will tell you these two groups are fighting today. There's an increase in violence between these two groups. What you want to know is, and maybe the news media will even tell you, they're fighting over this issue. What you want to know is, is this new? Do they usually fight over this issue? Is this something new? Do they usually fight in this area versus this area? What's the context about this? I work with a number of development groups where the interest is, they're going to go build a school in a certain area. Oftentimes, even with the US government, they'll send out teams to go build a new school – they know everything about the geology and the climate, they know everything so they can build a school that lasts 100 years, that building. The problem is what they don't know is this area, women and men aren't allowed to be educated together. So they build that wonderful school, it's amazing, and then it turns out it's not very useful to us. Or it's built in a certain area where maybe this ethnic group is not allowed to enter this area.

The problem is that type of information, like Libya, for example, or Syria, all these areas where you have – Libya, in particular – all these ethnic, tribal, other connections. These are relatively foreign, especially to

American and western analysts. I've done a lot of work on mega-cities, which are of great interest right now on the US side. A lot of people say mega-cities are so foreign, they're so different, how do we understand them? Tokyo is a mega-city. Tokyo was the very first city actually studied, when the US open source intelligence community was formed in the 1930s, Tokyo was target one. Moscow. Incredible amounts of data behind – well, Moscow is a mega-city. But Tokyo. But things like Lagos or Dhaka – there's nothing really different about this except in the case of Dhaka, there are 1,251 articles in the Bengali language translated by the US intelligence community and the open source community in a 22-year period, I believe, or a 20 or 22-year period. There is more coverage that they monitor out of, I believe it's Russia, each day than out of all Latin America, Spain and Portugal combined.

So part of this is pivoting and realizing that a lot of the things where we think about a lot of parts of the world, it's not that these parts of the world are something fundamentally different. It's often because it's a different culture or a different language. Really, if you invest the time in understanding this culture or this language, you have a great understanding of what's happening there. I know that's a little bit of a tangent from what you asked, but yeah, so, anyway.

Question 5

You made reference to the *New York Times* and the BBC, for instance, and obviously driving narrative and so forth. The BBC is not worlds apart from the *New York Times*. So how do you adjust your algorithms, your analysis, when it comes to offsetting the outlets that establish this narrative? Obviously, alternative news media, if you're evaluating volume, is going to be dwarfed. So how do you approach that?

Question 6

You seem to be doing such amazing things and seeing what you can do at the moment, but in the future are you going to be looking at what you can help solve? What problems you can tackle? Or are you just going to kind of combine all these amazing datasets and then put it out there and let other people do that?

Question 7

My question was about propaganda war. You started with saying any information is biased, but information is more or less biased. Is there something that your analysis can do about it, about big data?

Kalev Hannes Leetaru

I'll answer these two questions together, because they're kind of related. One of the things that people ask me is, how do you decide what sources to monitor around the world? I sit down with five Russian experts and I'll get five wildly different answers as to what's the authoritative media in Russia. Really, to me, that's a false question, because it depends on what are you asking. For example, a question might be with Russia Today, with the Ukraine conflict – what was it, one or two news anchors resigned over their coverage. You might say, did that impact anything? Did Russia Today course-correct or was that inconsequential to them? Russia Today, is that really media? Is it more a propaganda arm?

So a good example of this is oftentimes you want to understand what is the narrative within something. So in China's case, for example, North Korea – remember when the story came out that he threw his uncle to the dogs, feeding alive. That came out through the *South China Post*, it was one outlet basically. That's a good example where being able to say this came out from this one outlet, it's not in any of this other coverage and this outlet is known to kind of sometimes be used for telegraphing, that's a very powerful

way of being able to see what outlets is it going. Or in the case of the Malaysian jetliner, when that went down, being able to say two and a half weeks later, I think it was two or three Russian media were still running the story that it was a botched assassination attempt of Putin. That can be very powerful because there you know the narrative. It's highly unlikely it was a botched assassination attempt on Putin, but what's useful is to know that that narrative is still existing in this tier of media. In Russia, this stuff doesn't happen without some influence or interaction. So it's giving you some interesting pictures.

So the perspective here is basically try to grab everything that someone might say is potentially media. That can include the political blogosphere. Part of what I've been doing is going country by country, to countries that I have connections to or that other colleagues have connections to, NGOs that operate in these countries – basically, anything that I can get that is essentially a catalogue of the media in that country. Or what an average person might think of as media. That gets to that whole question of what's a blog versus news and all this wonderful stuff. Anything that could possibly determine that.

Then you can filter that down. You can say, for example, what's the views in Israeli media of what's happening in this country? What's the view through Russian media of this? What's the view through tier one, tier two? What's the view of the political sphere? So being able to slice and dice that. Of course, you do have issues. North Korea, there's just a lot less access to that media and a lot less useful media in North Korea. In Russia, even the stuff that's state-controlled is actually useful for understanding what the Kremlin is trying to force-project. In North Korea, they've never really mastered that. Or like Vanuatu, there's simply less media there than in Russia.

So this does require a lot of careful normalization. When you think about working with this media, and certainly bias – this is why people can do different slices and dices through the data. For example, I've had people that have come and said: here's my catalogue, what I consider to be state-controlled media in Russia. That scholar or that policy person or that government analyst or that whoever has their list. You can slice that data down and say what's the view of, for example, in Syria – what you saw a lot in the Russian media for a while, of course, was when the government was winning, that was actually reported. When the rebels won against the government, that stuff didn't make it in a lot of Russian press. Some of it did but there was certainly less of that. Or Nigeria, for example – I haven't looked post-election. Pre-election, news coverage that portrayed the government as losing against Boko Haram had a very short shelf life. A lot of that material, the newspaper would put something up, it would just vanish. So we also have a partnership with the Internet Archive where we send them a master list of all the URLs for online media – broadcast and print are handled differently but online media, we send a copy of every URL that we find worldwide to them, and they permanently archive it in their archives collection for preservation.

So bias is a fascinating question. That's why a lot of people slice and dice through it. Hopefully that answers a little bit.

Then to your question, yes. GDELT is used in a lot of active places. So right now we do a custom feed that's used with UN OCHA. We filter down anything that has to do with active disasters. Anything that UN ReliefWeb has basically catalogued as an active disaster right now, we filter that and provide that list for them. Then that gets deployed out to things like Micro Mappers and some of these other programmes where humans then review all that coverage. Or for example, when Ruby went through, what we did actually is there was an interesting catalogue of aid commitments around the world. So major NGOs coordinate with each other when they do aid shipments, but you think about some church group decides to bundle up maybe \$50,000 of medical supplies and just mail it somewhere. This stuff happens. How do you basically get all this? Or even reputable NGOs, again, there's a chaotic-ness that goes with these disaster situations.

So one of the things we're doing is looking worldwide, anyone that was discussing aid commitments into an area. Dollar amount, what was being committed and so on. So being able to flag that across these languages, across the world, build an update every 15 minutes, here's what we're seeing. Or even being able to flag what NGOs are even being mentioned in that. You might say, wow, World Vision is coming in over here, that's interesting. So that ability to – so basically the examples I give here are more of the analytic cases, because I think they're interesting. But yes, it is being widely used in a large number of situations. Many of those are public, many of those are NGOs or others don't necessarily advertise. There are NGOs, for example, that use this for security purposes, to say we've got deployments in areas that we don't necessarily want the local – we work with the local groups, we don't necessarily want the central government that we're in this area, because they might come and kick us out. So how do we watch what's happening in that area? How do we watch movements of things? Apparently, I don't know military tactics, but allegedly there are groups that have been using this to look at ISIL/ISIS/Daesh attacks, and actually look at areas where they're using traditional Iraqi tactics versus tactics that are not Iraqi tactics. So foreign fighters versus de-Baathification. Again, I'm not a military practitioner, so I have no idea how that works.

So the short answer is, it's used in a lot of cases. As with any open dataset, some of those are things that people splash out there and they advertise to the world. A lot of them are not necessarily applications that people splash to the world. That's one of the interesting things about an open dataset. The goal of this is to enable fascinating research on things. I'll see some academic paper from some group and be like, wow, they're applying some Bayesian something or something, and they have map equations that are 10 pages long, and I have no idea what this equation does – but this is really cool what they're doing with it.

That, to me, is one of the exciting parts, because when you have a truly open dataset that's available for anything, freely available to the world, what you get – you think about some NGO, for example, that's doing work on the ground, they don't have data statisticians there doing neural network, Bayesian network analysis, such and such algorithm. That's not what an NGO specializes in. You get, for example, some academic or some Wall Street group that publishes papers saying here's how you can do it, and you can do, for example, impulse dynamics that catch these sudden surges of things. All of a sudden, that gets useful for all kinds of purposes. This ability to really force leverage – shortly we're going to have some additional ability for mapping all this stuff. Anyway, talk to me afterwards.

Thank you so much, this has been fantastic.

Caroline Baylon

Thank you so much. Sorry to cut you off but you've given us a lot to think about. Thank you for a great presentation.