FROM BIG DATA TO GLOBAL DIPLOMACY TODAY AND TOMORROW

INSTITUTE FOR THE STUDY OF DIPLOMACY

MASTER OF SCIENCE IN FOREIGN SERVICE

YAHOO! FUND ON COMMUNICATION TECHNOLOGY, INTERNATIONAL VALUES, AND THE GLOBAL INTERNET
The Yahoo! Fellow in Residence explores the relationship between new communications technologies and international values in varying national and international contexts. Projects should address the issue of applying international values and principles (e.g., free expression, regulation, citizenship and privacy) to new communications technologies and the diffusion of information in rapidly changing global environments. These might include large and evolving economies—including Brazil, Russia, India, and China—as well as interactions between the developing and developed worlds. Projects can draw on insights from many disciplines, including politics, economics, business, and socio-cultural research.

It is expected that the Yahoo! Fellow will undertake research and writing, some of which will be posted on the website of the Institute for the Study of Diplomacy, during his or her stay. In addition, the Fellow must organize at least one workshop or conference during each of the two semesters in residence, work with graduate fellows supported by the same Yahoo! grant, engage in co-curricular offerings and other interactions with the Georgetown student body, and participate in the professional life of Washington, D.C.

The Yahoo! Fellow in Residence will be a part of the Institute for the Study of Diplomacy (ISD) at Georgetown University. The Institute will provide office space, supplies, and administrative support during the fellowship year. The Yahoo! Fellow will have the opportunity to be a member of a group of ISD associates and fellows drawn from the United States and foreign government agencies.

Within the School of Foreign Service, the Master of Science in Foreign Service Program (MSFS) will provide a locus for the applied curricular and research support activities associated with the Yahoo! Fellows. The Program’s multi-disciplinary faculty, including experienced public and private sector practitioners, comprises a broad resource base for research and collaboration on globalization forces and impacts across an array of diverse sectors and cross-national applications. Relevant results generated by the Yahoo! Fellows’ projects could be incorporated into the MSFS Program’s curriculum through guest lectures, special seminars, case studies and/or course modules. It is anticipated that research projects may thereby yield multiplier educational effects that extend beyond the tenure of a particular Yahoo! Fellow. MSFS graduate students will also be selected to receive partial-tuition fellowships as Junior Yahoo! Fellows. The selected students will engage in study and research associated with the Yahoo! Fellow in Residence and/or other faculty engaged in related project activities.
From Big Data to Global Diplomacy

Today and Tomorrow

2014 YAHOO! CONFERENCE
YAHOO! FUND ON COMMUNICATION TECHNOLOGY, INTERNATIONAL VALUES, AND THE GLOBAL INTERNET

A transcript of the
April 2, 2014 conference
The online world is allowing faster, more representative, and round-the-clock access to societal behavior around the globe. Today, people from Bangladesh to Buenos Aires busily tell one another and their neighbors what they see, what they think, and what is important to them, offering unparalleled visibility in the digital heartbeat of global society. The proliferation of always-connected mobile devices means citizens and participants are often the first to report on emerging events, streaming photographs, videos, and ground reports as events unfold, while social platforms have become one of the primary organizing tools for rebel and opposition movements. Moreover, the constant stream of daily life that flows across online media platforms provides rich contextual background information on the narratives of each region and culture.

A growing chorus of voices argues that “big data” is a usable, practical tool to help foreign policy makers, implementers and analysts increase their understanding of global issues. For example, computerized analysis of massive amounts of world press reporting yielded an innovative map of Ukraine’s protest violence useful to any individual grappling with the present and anticipating the future of that region. Can similar analysis of the “thoughts, discussions, and debates expressed in public social conversation” provide similar insights for foreign policy practitioners, scholars and students?

Georgetown recognizes the Yahoo! Fund on Communication Technology, International Values, and the Global Internet for making this program and the fellowship possible.
Contents

Introduction 1
ANTHONY AREND
Director, Master of Science in Foreign Service Program

The State of Big Data 3
KALEV LEETARU
Yahoo! Fellow in Residence, 2013–2014

The Social Media Landscape 7
WILL MAYO
GNIP

Big Data, but Not for Superman: Engaging and Empowering Grass Roots Peacebuilders 11
NOEL DICKOVER
U.S. Institute of Peace

Big Data to Visualize, Analyze, and Forecast Global Issues 23
KALEV LEETARU

Panel Discussion 43

About the Speakers 51
Introduction

ANTHONY AREND
Director, Master of Science in Foreign Service
Edmund A. Walsh School of Foreign Service

On behalf of the Institute for the Study of Diplomacy and the Master of Science in Foreign Service program, I want to officially welcome everyone to Georgetown University for our conference today on big data and global diplomacy.

We have quite a great group of presenters today, and our impresario is Kalev Leetaru, our Yahoo! Fellow at Georgetown this year. Georgetown, as you may know, has been greatly, greatly honored to have Yahoo! establish this fellowship over the past several years. We hope it continues indefinitely into the future.

Kalev has assembled a wonderful panel of participants to talk about the issues before us. So, without any further ado, I am going to turn the podium over to Kalev, and thank you very much for your energy and all your hard work on this.
We are going to focus on something that I think is a very important area of things today, which is essentially big data and diplomacy. How is big data really reshaping how we think about the world and what the future of public policy can be?

I want to start off with a quick story. A story that begins four and a half billion years ago with the creation of this amazing planet called Earth. About a billion years go by till the first life starts wandering around on this primordial planet. A few more million years go by, till about 50,000 years ago. The first modern humans start walking around. Something very interesting happens about 38,000 years ago. Those early humans start leaving a permanent record. They start leaving cave paintings. About 8,000 years ago the first surviving evidence of writing begins to appear. Five hundred years ago the Gutenberg Bible show up; Gutenberg, with his printing press, can start mass producing information. About 20 years ago the modern web browser shows up; ten years ago, arguably, social media shows up; each bringing with it an exponential rise in the ability of human society to document itself, to tell us what is happening.

Think about it: we know more today about your average college student; we have a better documentary of their Friday night than we do of some societies from 2,000 years ago. You think about what this means in terms of documentation. You think about the average citizen. You think about society around the world today. We know so much about what is happening within those societies. People are able to document what happens. If police, for example, beat someone up on the street, today chances are that someone might actually be filming it. They might
upload that to YouTube. There’s actually a chance that people around the world can see that. You think about what this means in terms of our ability to really capture human society.

If you fast forward to today, we have reached an incredible point. Electricity is reaching the farthest extents of the world, and with electricity comes a connection to the global network. This is a really interesting image [slide]. It may be a little bit hard to see, but the red dots here are locations of where there was one or more tweets. So this is, I think, November of 2012. And each red dot is where there were one or more tweets. Blue dots are areas where there’s more electricity than Twitter. And you can see that pretty much there’s not a whole lot of areas that are bright blue, other than sort of rural Russia, and of course China, where Twitter is officially banned.

What is really interesting about this are the white and the red areas—white shows areas where there is as much power as Twitter—Twitter, really—where there is power there is Twitter. And that is really amazing today, that even if we are talking about areas like Africa and Latin America, Twitter and other social media forms are really connecting the world. They are allowing us to hear from people. They are giving us just this enormous volume, this rich volume, of information.

Facebook alone claims almost a billion members connected by over a trillion connections. Now obviously a lot of those billion accounts are someone’s cat or the Statue of Liberty. This is something [slide] that Facebook created in December 2010. What is amazing about this is that you could literally make a network diagram like this, that a company could basically run some algorithms and make a network of how the world is connected. Again, according to Facebook, while it is a highly biased view of the world, the ability exists to construct a network diagram of human society.

If you think about our digital world today, a third of the population is online. Today there are as many cell phones as there are people on earth. Think about what this means. Facebook alone claims 240 billion
photographs. The amount of material that is out there today just blows my mind. Every year, 1.6 million days of video are uploaded to YouTube, every single year. You think about that. That amount of just raw material is out there.

Every day and a half—I am sure these numbers have increased—there are as many words posted to Twitter as the entire New York Times over the last half century. What does that mean? Now, again, a lot of this is potentially automated bots or advertising or other things, this live documentary of society, every minute, tells us something about human society.

What is amazing here is that this data revolution is coming alongside a super computing revolution. We have the machines now. What is fascinating is that this analog of the heartbeat of society is increasingly being digitized live. People are photographing it. They're blogging it. They are essentially live digitizing human society. This means our analog and our digital heartbeats are increasingly beating as one. That is very fascinating; it means that machines can monitor, map, and model this data. If ever we were at a point where we could use machines to create a virtual heartbeat of society, it is fast approaching.

Today, people from Bangladesh to Buenos Aires busily tell one another and their neighbors what they see, what they think, what is important to them. The first confirmation of Khadafi's capture was a cell phone video, captured by participants. Osama bin Laden's super-secret takedown was live-tweeted by a Pakistani journalist saying, “There's a foreign helicopter flying over.” It was confirmed here in the United States, not by an official statement from the Rose Garden, but by a former administration official tweeting, “I heard from my buddy in the White House. We got him.” Think about what that means, today, in terms of society.

This is one of my favorite images. This is the election of the new pope in 2005. You notice there is one little camera phone down here. Compare this with the image of 2013—I love the iPads that has been used as cameras. If you think about this, in 2005 the only views we had of this was the government’s view. In 2013, what we are seeing now is every one proba-
bly live-tweeting, live-blogging, and live-Instagram-
ing what is happening. It means we have all these
different perspectives on what is happening around
the world.

As of 2010 nearly half of the news media around
the world—broadcast, print, et cetera—was avail-
able through electronic means. There was an elec-
tronic version of it available. This is very critical
because what this means is that for the first time ma-
chines can actually process all this stuff. This is not
low-power broadcast from an embassy somewhere;
that is not something that machines can easily pro-
cess. But when it is on the web, machines can access
it. They can process it. They can make sense of that.

The social media world is this enormous, enor-
mous entity. To talk more about that, we have Will
Mayo from Gnip, which is one of the major social
media companies that provides the data that is really
the lifeblood of our understanding the world.
The Social Media Landscape

WILL MAYO
Gnip

World's largest and most trusted provider of social data.

A few facts...

- Customers in 41 countries including many of the world's largest software companies
- First data partner for Twitter, Tumblr, Wordpress, Foursquare, Disqus and others...
- Only company with access to the complete archive of Twitter data
- Distribution network touches over 95% of the Fortune 500

Critical layer in the social data ecosystem
Unrivaled access to key publishers

These data sources provide amazing insight into global behavior

Social data is great when used for alerting...

- Content is virally disseminated and has been known to travel faster than traditional news
- Data can be used as an alerting mechanism
- Data has a multi-media context that provides on-the-ground insight into situations
Providing unique points of view for diplomacy

- Unique insights gleaned from viewing specific account activity during Hamas-Israel crisis
- Analysis performed on followers of accounts
- Data demonstrated how information was spreading through followers and retweets

Sources offer global perspective...

- Twitter itself has a large amount of international content (78% outside US)
- Blogging and commenting platforms (Wordpress, Disqus) are also used worldwide
- New sources like VK offer insight into specific regions

How to gain insight through social media data...

Turning social media data into structured data

Full coverage of historical and realtime data
Big Data, but Not for Superman: Engaging and Empowering Grass Roots Peacebuilders

Noel Dickover
U.S. Institute of Peace

I am from the U.S. Institute of Peace. We have recently created the PeaceTech Initiative, which I will briefly talk about. This will lead to the bigger issue with these wonderful data sources that these two folks, and others, have. Who are we making it for? Who is actually seeing it? What are the implications of its use? I am really talking about big data in terms of its use in conflict zones.

Our world of peacebuilding is changing dramatically and is based on changes in access to and use of technology. I think you all probably get this better than anybody else, but the shift is continuing. The institute has made this investment in a project called the PeaceTech Lab, which will be coming out this year. Basically we are looking to use media, technology, and data to support and cultivate grassroots peacebuilders and give them the information that we hope will help them actually work in conflict zones.

We are doing a number of things. We do media programs in countries, so we have a reality TV show in Iraq. We have radio programs in Afghanistan with curriculum-embedded content. We do these exchanges between local technologists and civil society organizations. I developed a program at the Department of State for the last three years called Tech Camps, which we have done now I guess in over thirty countries, and we are bringing this toward peacebuilders; we have done a series of three. We just finished it in Iraq, and I could talk about that
The PeaceTech Lab

Collaborating at the intersection of technology, media, and data to devise effective means of reducing violent conflict around the globe.

CONVENE

Bringing together experts across disciplines and across sectors to address real needs on the ground

CONNECT

Connecting our work to ground truth from the field via the Lab’s Open Situation Room Exchange

BUILD

Designing new, low-cost technologies appropriate for the environments where peacebuilders operate

INSPIRE

Inspiring networks of local and global technology-empowered peacebuilders to develop tools for social good
if you all are interested. How many of you are familiar with the “Blogs and Bullets” program? We just came out with a really cool study of the impact of social media in Syria over time.

THE BIG QUESTIONS

The last one is the one I really want to talk about. This is our first time kicking the tires on this so if it is a little incoherent, hopefully we get there. This program is called the Open Situation Room Exchange. It starts with a big question. Who are we building the data for?

You know of the notion of conflict early warning systems. “If I get this alert and I tell Obama, he is going to fix Nigeria.” Right? That is going to happen. Right? How many of us believe this? How many times have we seen that behavior? There are massive efforts to inform really smart people, far away from conflict zones, but who is going to actually make a difference in the conflict zone?

If we are talking about changing the attitudes and beliefs of the people there to choose nonviolent means for resolving conflicts, is this something Superman can help with? Do the arrival of UN peacekeepers really make a difference? And if we are gearing it for these folks, what does it look like? How are we doing that? This is not an easy problem. I mean, this is a really hard problem to think through.

I am trying to come up with a term for data that folks use on the ground. I use “local data.” Does anybody have a better term for that? I am not talking firehoses of Twitter doing stuff, but actual real people figuring out how to do their agriculture project or whatever that is. That is often what they are relying on. There are some challenges with that. So everybody that’s working in these bases has data, but they are not really releasing it.

For example, the Nigeria Mercy Corps has this wonderful approach of identifying conflicts and sending in a mediator to go off and do something about it. But who has access to that? It is really the folks that Mercy Corps trusts. Which leads to another question: How do you build that trust over time?
This is a real challenge, and this is one of the reasons we like to bring together local technologists and local peacebuilders, because the technologists in the country are going to know the profile: Do the cell phones search, are they GPS-enabled [Global Positioning System¹], is there an SMS gateway?² I mean, with all the things that you can think about, actually thinking through how somebody releases data into a public environment is a really scary thing. If I am not a data expert, and I know the data contains private information, then I also know I have information that may be risky. How do I just even think through that process if I am not a data expert?

The peacebuilding field is not what most would call data fluent. Most people in this city are not data fluent. So imagine you are working with folks on the ground who are really dedicated, doing good work, do they really understand how to take the GDELT³ database query and get information that is going to help them make better decisions tomorrow?

And data access. I mean, this is a big problem. I was in Burma in January [2014]. They have a four-

---

¹. “The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites.” Source: Wikipedia, http://en.wikipedia.org/wiki/Global_Positioning_System, April 15, 2014.


³. “The Global Database of Events, Language, and Tone (GDELT) is an initiative to construct a catalog of human societal-scale behavior and beliefs across all countries of the world, connecting every person, organization, location, count, theme, news source, and event across the planet into a single massive network that captures what’s happening around the world, what its context is and who’s involved, and how the world is feeling about it, every single day.” Source: Wikipedia, http://gdeltproject.org/, April 15, 2014.
teen gigabyte pipe for the country, outside of the military. They are not using video. Right? They are not. It is a very small piece that even has cell phones. So this becomes a critical problem. How are they accessing this data? And who is looking at them as they access the data? So this is a real set of challenges that we are testing. We are trying. I am not going to come forward and say we have the right answer and know how to do it. We are much more in an experimental phase here.

The idea we have come up with is to see if we can build platforms that help connect people working in these conflict zones. There are a couple of design principles. We want this as open as possible. But again, based on the trust environment, it is really a matter of can we find ways that these people can use to communicate and connect. It may be in a protected conversation; it may be we release data that comes out of that conversation for others to use.

The U.S. Institute of Peace was founded by the federal government. NSA [National Security Agency] is funded by the federal government. If we did a proprietary platform, what is going to be the fear of the locals? The use of open source tools, therefore, is fairly critical to us. If we ever build modules, the idea is to release those back to the community while reducing concerns for its use. Over time we really want these exchanges in conflict zones to be locally owned and managed by the technologists and peace-builders there. So that is the approach.

I do these sort of weird networky diagram kind of things. I am sort of a—I do not know if most of you know, but I carve pumpkins. In October, BBC did a special on me to put around the world. So I am sort of always having to do similar things like this, so apologies for the sort of weird view.

So imagine we have this Open Situation Room Exchange, this platform. We have a data store associated with it, and I would love ideas for how we want to build that. There are a bunch of different options.

- We can think through a thematic view of open situation rooms, like how do you get a civic mobilization activist to use technology more effectively. That could be one conversation. And perhaps by not
talking about it in the context of Libya, but just really advancing the state of the practice.

- You can have a regional use looking at the Sahel region or a country.
- And then, where we want to focus right now is really the situation. So not as much giving a letter grade to a country as actually focusing on an issue in a country. How many in this audience have worked with Nigeria, for instance? If I’m talking about Boko Haram\(^4\) in northern Nigeria, that is a very specific problem, right? We have a former ambassador whose fear is in five years northern Nigeria is going to look like northern Pakistan. How do we think through that problem? What is the data that supports that? How are we going to do that?

We have the international development community coming in place. We really want to connect primarily with local peacebuilders, though, in the local tech community; and we have folks like you. All of these people engage in this conversation. What we really want to sort of tie this together is what we call “Peace Tech Fellows.” The idea is to find folks in the country, technologists, bring them back to the Institute, spend six months working through how do you use data, how does this whole process work, train them, send them back and actually have them run the local situation room. If it is locally owned, locally managed, the hope is that we are going to start thinking through this mix of big data and local data.

Obviously, there are a number of different ways to tie things in. Again, we are sort of in the proof of concept phase here. We believe the main idea that trusted conversations will lead to better situational awareness. You are going to notice gaps in the data. You are going to find really cool, awesome datasets

---

that people are using, and go through a process of trying to release that to public datasets that everybody can access. And that is really the thought. Over time, just working with Kalev over the last couple of months, thinking through how do these big data stuff can really help somebody on the ground tactically think through what they are going to do tomorrow. We are not there, but we are certainly thinking through that process.

What are the conversations we are talking about here?

**Stress data.** There is a whole variety of things that impact conflict zones. There is a case being made that the reason Egyptian President Mohamed Morsi was toppled was due to massive flour and gas shortages. People were pissed off. Then they started caring about the politics. But if there was adequate flour and gas, and people did not have to stand in line to get gas and not get there after three or four hours to get their bread in the morning, maybe it would not have gone to the extent that it did.

The idea is for people to develop scenarios of what they think might happen. That has an expectation of where we believe this will go. What you want to do is get this conversation around potential mitigation strategies, in thinking through what are the actions that are going to really make a difference. This is really critical for the international players. I think we have all encountered fairly simplistic views of if you throw money at a problem, it will lead to solutions. Try working with civil society in Iraq. They have had more money dumped on them than anybody else in the last ten years, and it has not raised their capacity at all. It has made a lot of people very rich, but you do not have an active civil society that knows the problems and thinks through them. Instead, you have large NGOs who sort of have a bucket. They go around and say, “What is it you need me to do? I will do that very well. I am trained to help internationals do what they want. I am not thinking through what are the problems of my country that I need to fix.” In some cases, that is why you end up there.

There is also this tendency in the development world that everybody has their own network on the
ground. “Well, we have invested in our network, we talk to them.” People do not own networks. People belong to networks. So the idea is how do we connect all the folks on the ground who are really working this, regardless of who they are affiliated with. How do you connect it? The idea is to have more of a commons approach to this, where people can come in, access data, exchange it, but it is not really owned in that sense, other than by the folks on the ground. And ideally, we want this to turn into a public utility for peacebuilders to really think through what their strategies are.

So what can big data do for me? I need to find technologists in these places. I have gone and worked with technologists in 20-some countries. Some of them, like Guatemala, are amazingly connected. Everybody knows everybody. It functions well. If you have had lots of BarCamps, that sort of is a good indication.

You look at Honduras right next door, and it is basically individual technologists working on contracts to U.S. companies. They do not get together for a variety of reasons. But if you bring them together, they love it. And if you give them an opportunity to work on social good projects, usually it is the first time they have ever done it. And they really like being Superman. They like being the person that can help those on the ground. Although everybody’s told me the real analogy should be Batman, because they do not have super powers, but their genius IQ and ability to work is what they are doing.

But how can we find these folks? Are there some innovative ways to see what technologies are being purchased or how can we connect with technologists to expose them to the idea of working in the

What Can Big Data do for the OSRX?

- Help me find & connect w/technologists in conflict zones
  - How are they communicating?
  - What are their skills, gaps?
- Better ways to share stressor data
- What are the data gaps in each conflict zone?
  - How can we “infer” answers from big data, or foster crowdsourced approaches to fill these?
- Automated Yelp for Peacebuilding?

---

5. “BarCamp is an international network of user-generated unconferences primarily focused around technology and the web. They are open, participatory workshop-events, the content of which is provided by participants. The first BarCamps focused on early-stage web applications, and were related to open source technologies, social software, and open data formats.” Source: Wikipedia, [http://en.wikipedia.org/wiki/BarCamp](http://en.wikipedia.org/wiki/BarCamp), April 29, 2014.
social goods space? How can we share stressor data better? How does somebody with a simple feature phone with access to all those databases find a useful way to think through what he is doing? Maybe he has the answer to that, but that is a tough problem.

And where are the gaps in data? What are we really missing? If we start thinking this is what election violence looks like in these five countries, and we have an election in Nigeria coming up next year, or in Burma, how can we start thinking through this to say, “Well, we really need this piece of information if we are going to start countering hate speech online,” or whatever the issue is. How can we start thinking through the data gaps and either use crowdsourcing techniques or whatever else to find them.

My hope and desire is for peacebuilding. There has been so many attempts of, “Everybody, tell us what you are doing in Libya. Now we know what everybody’s doing.” And we are not going to update it very well, are we? So any time it is like this, “Give me information by hand that shows up in my database,” it is not going to work. Millions of dollars have been spent because it is a really important problem, with many challenges. Do I really want to tell you where my people are in Libya? Because there may be people that want to know that who do not have their interests at heart. There are a lot of problems associated with it, but if you are going to start connecting the peacebuilding community, you are going to need to sort out this piece. And it is not a manual problem. It is a big data problem.

So here is where we are. We are building a platform in Triple 7. It is hopefully going to be ready for testing by the beginning of May [2014]. We are working on partnerships with some of these data providers to figure out how their stuff can automatically come in, and when Apollo’s system updates every fifteen minutes you would love it if there was some impact in Nigeria for the folks working there, that they see that in a way that is meaningful for them. This is a challenge because the stuff we care about in Nigeria, “Wow, there are fifteen attacks that happened today,” is not something somebody in the delta region is going to care about. And what are they
going to do with that information. I mean, we really need to work with them to figure that out.

We know there are mobile connectivity options. FrontlineSMS is probably what we are looking at, but that is still something to be reached.

Finally, we are actually looking at a physical instance of this here at the institute. The idea is to get teams of folks from different organizations to interact with the data, start doing what—if scenarios, connect with the folks on the ground, and hope, especially for the policy makers, to give them a sense of the complexity they are talking about. This is really not something you can wave a magic wand and dump buckets of money and expect a to solve the problem. There is only so much money any country can take in a period of time. If Congress tells USAID, “You have got to spend X amount of money in Afghanistan in the next year,” that does not mean Afghanistan is going to be that much better. I mean, there is just so much you can do from that perspective.

QUESTIONS AND ANSWERS

Question: The comment is, when you look at the index card—not that people have these anymore in libraries, but it used to be an index card in a library had content on it that was not necessarily—the terms that were used on the card were not the terms that appeared in the book that it was describing. It was often one level of meaning higher than the content itself. Often, a lot of these kind of data, “if we just get all the data, then we’ll have the visibility we need.” And there is a level of sense making and of meaning, I think, that is not just plug in the data. You get the knowledge you need.

So there is, perhaps, as much attention to setting up the human conversations around the data, the knowledge sources, as there is getting technology that can put all the data bits together. That is the context of the question.

Last year, some of us ran an experiment that was about shaping the conversations around, in this case, what would the future decision support needs of field practitioners look like in three years’ time? It was a pro bono experiment, with Gateway. It went a certain degree; some things were dubbed beautifully and other things we could have gone further.

We would love to take that forward. But the question is about the human dynamics of getting it right.

Dickover: So if I just go back to Burma, for example, in August the whole IT profile of the country is going to change. Telenor and Ooredoo are two international companies. Telenor is setting up a 2G network; Ooredoo is setting up a 3G network. Currently, sim cards are somewhere between $50.00 to $100.00, which is much cheaper than they were. So anybody that purchases them will also purchase a Smartphone. Approximately 90 percent of the phones in Burma are Smartphones. That is a weird profile of the country, right? So it is almost like in the ‘90s how AOL was? That’s what it is with Facebook there now.

MEDO is an ICT group, and they did this session—they have the biggest BarCamps in the world. Six thousand people show up for their BarCamps. I think, based on spending some time there, it is because the education is so bad that any opportunity for folks to learn they will come together and do it. But, so in August you are going to have sim cards drop down to $5.00 apiece. You are going to have a massive expansion—we hope—of the pipeline into the country. What’s that going to do to the dynamics?

This is really a problem for peacebuilders, because they never look at the technology. They are conflict resolution folks; and it is amazing to watch them. I went with two of them there. They have a five to seven year timeline of building trust. Once they have the trust, then they plant doubt, and then they start to move forward. So we are not addressing the core issue there because we do not have trust with the police department or with the general accounting when working with civil society. The world is going to shift dramatically.

Hate speech is already a problem on Facebook today. What about when you jump from 7 percent to 47 percent of the population on cell phones? You
have sixteen armed conflicts in the country. There are notionally fourteen peace agreements between them. Yes, it is an authoritarian regime, but it is not controlling in the way that you might think. Our focus is really thinking through hate speech problems, especially hate speech leading to violence before an election. How do you pre-position and get out in front of that?

The problem with a lot of these data things, if you read peace-building assessments, is that it is steeped in this rich dialogue that describes the dynamics. At USIP, we have some of the most amazing experts on these countries. Back to the Morrissey thing, when that hit, the conversations inside USIP were so much more detailed than anything you could find online. It is people that have met with everybody there and talking about what is happening.

So if you are just thinking through the data piece, you are right. You are going to miss something. When suddenly everything turns into something new, the sets of relationships you thought were in place may not be there anymore. How do you recognize the change has taken place? How do you think through the dynamics

What this gets peacebuilders to do is follow the principle of do no harm. They do not really want to try anything unless they can think it through. I used to think that diplomats at the State Department were the worst about not taking an action before they knew what was going to be the result. They have nothing on peacebuilders.

So here I come, talking about crowdsourcing. Can you control that? If I get Burmese youth to start crowdsourcing the monitoring of hate speech online and swarming the results, they may say some nasty things. Is it my fault that they have done it? Because of that, should I not take any action? It’s a real challenge to the peacebuilding field. It is a different way of thinking about the problem. It is a very uncomfortable problem, because you do not want to be the one responsible for making that change.

Question: Have you thought about the ethics of collecting this information? I think in an authoritarian regime it is almost certain that you are not going to be able to keep it secure, especially if they have any sort of cyber command.

Dickover: That is a huge concern. In some places, if all you have is mobile, and the government owns the telecom, it is not going to be secure. This is why local technologists are so critical. They are going to tell you what the profile is in this country. If you connect in with civic mobilization folks, in that sense you really start thinking through the level of engagement in terms of a calculated risk. So every activist in a conflict zone is taking a risk on a daily basis. The question is, do they know it?

A great example was an exchange in Mexico around women empowerment. Leading up to it, two activists on Facebook who were just doing some amazing things in the Chihuahua region were killed. They were gone. And nobody called up and nobody cared. It seemed that they did not realize what they were doing. Their process had put them at risk. It is a very different thing if you know you’re taking a risk.

From our perspective, it is making them aware of those risks. It is a concern when we release data. There is no way that you are going to keep out the bad elements from open data. They are there. They are on social media.

This is another debate point. If the same information can be used for harm, should we not create it? I would argue the same information is already being created for harm. That is critical both for what we do and how to safely interact online. In most of the tech camps I have done, that is a key component.

Question: What does the Peace Lab do as far as dealing with these multilingual landscapes?

Dickover: We do have one thing in our favor. The language of technologists across the world is English. Technology is this really interesting thing where if you are in Honduras, and a technologist, your chance of ever getting on the international stage is close to nil. But social good gives you an opportunity to get to 92Y in New York and a whole series of other things. So it really impacts their ability, both the money they receive locally and stature, which is
really important to technologists. I think there are some real incentives on their part to do it.

In working with technologists, we do have a leg up, but frankly most of the really important conversations are going to be in the language of the country. So like the PeaceTech exchanges that we do, I have done them with up to three languages being used. So the plenary sessions will have translation booths with people with headsets. The majority of the event is all small group discussions.

In some discussions, we train twelve different technologies over an 80-minute period. For example, Samantha Barry from BBC has come to a number of these. She will show people how to take their cell phone, take a video, edit it, and upload it to YouTube. She could do that for anywhere from five to twelve people in an 80-minute session.

You have others that are showing people how to take Mapbox and map data effectively? Or, how do you do digital storytelling? It is like a BarCamp with structure. The idea is to find this local peacebuilder, who connects with the technology and a technologist, and then the light bubble will go off. It will identify problems that they can start working on.

During the first day, the technologists are leading most of the sessions. It flips on the second day, where the NGOs take the lead. Then the technologists decide which problem they want to work to help and address.

What I am able to do at the institute that I was not able to do at the State Department is to issue micro awards to some of the projects that come out of these workshops. If I give them $5,000, if they do something amazing with it, maybe we can give them some more.

I am not going to micromanage the 5,000 members of the group. If anything goes bad, okay, it happens. Again, our goal is not to spread this information evenly to everywhere. It is to identify and cultivate local leaders, local peacebuilders, who really have that sense of what is wrong with their country and where they think they need to go, and then we want to find ways to support them.

Our goal on the ground is to cultivate these people, support them, and connect them with the resources that we think will be effective.
I am going to talk about how we can use big data to try and understand questions that might interplay with the policy realm.

This is a collaboration I did with the Internet Archive. They have half a million hours of American television news, over the last three years. Very simply, the Internet Archive is essentially archiving American television news. One of the interesting questions raised from a policy perspective was what does American television news tell the American public about the rest of the world.

I think the latest Pew poll numbers shows that, for 70 percent of the American public, television still plays a key role for news. So the question was, when someone turns on the evening news what do they hear about? Is it just about the United States and Europe? Do they hear about Africa? Is it mostly about the Middle East?

To find the answer, we created a map with a time slider. The URL for the map is archive.org/tvgeo. For each day that goes by, it is displaying a dot on the screen where there was a mention of a topic in the news. It is looking for any mention of a location anywhere on earth and putting a dot on the map where that location was mentioned. What you can see from this is the United States gets a lot of coverage. Certain areas of Latin America are getting a fair amount of coverage. What is interesting in Africa is that you are getting some mention; South Africa is getting a lot of coverage; for the Congo or Angola, you are only getting mentions in a major city or two. To a degree, that makes sense.

One of the interesting things that we find with
American media can be taken from the New York Times. For example, the New York Times assumes that an American reader does not know much about Syria, does not know the geography of Syria. So when the Times reports on Syria it will write “northern Syria” or write a major city like Aleppo. The Times will not mention a specific hilltop, or a specific bridge, or a small province.

Essentially what this is telling us is that while the Middle East is getting a ton of coverage, that television news is not completely ignoring these areas, there is an uneven landscape.

From a policymaking perspective, if you are concerned with events in Nigeria, the unrest caused by Boko Haram, and you want to do something about it, ultimately you need to convince the American public that there is a compelling need here. You have to show that there is something for our attention.

That is one of the interesting things about data—the ability to take all this television news and make a map. In addition, if you go to the Internet Archive website and find this map, you will also find an alternative version. In this version, you can zoom in and compare the coverage of CNN versus Fox News versus others. You can actually drill in, click a dot, and see television clips.

This can be powerful. What data is allowing us to do is to really peer inside society at large. This next map was done last year [2013]: this presents one month of the Twitter Decahose¹ and color-codes each tweet by the language spoken in that tweet. This allows you to see the larger geography of the world. You can see, Russia. You can see just how much of Russia is not heavily populated. You can see roadways; you can actually zoom in and see the major transportation networks in these countries, see the hard boundaries between countries.

On further examination, it is quite interesting to

---

¹ Twitter sells commercial access to its streaming tweets. A decahose represents a package 10 percent of all feeds. Source: “Twitter Will Start Selling Half of All Tweets for $360,000/Year,” Lauren Dugan, (http://www.mediabistro.com/alttwitter/twitter-will-start-selling-half-of-all-tweets-for-360000year_b450), April 16, 2014.
see that language really does stop at national borders. You will see people on either side of the border, sort of isolated cases, but national borders tend to be pretty hard. In looking at the French border, French tends to stop at that border and it tends to be the other language on the other side.

Before we had social media, we had information in cyberspace. You could see a web page and have no idea where it was from, who authored it, or any information about it. With social media, the real issue today—the holy grail—is no longer what is being said out there, it is where it is being said.

If you want to know, for example, what is going on in Crimea right now, what you probably are not most interested in is a tweet by someone in Waco, Texas, “Boy, imagine what’s happening in Crimea.” You ideally would love a tweet from someone in Crimea, tweeting what is happening. Furthermore, reporters have gone to some of these areas looking for tanks and finding none. The locals would say, “What tanks? There have not been any tanks here. We don’t know of any tanks.” But when you go back to social media, you see that people have actually taken photographs of tanks parked here a few days prior.

People are really a social sensor network for us. Yes, we have satellites up above, but satellites can only do so much. The citizens on the ground provide the ability to document what is happening within national boundaries.

Other groups have used social media to look within a city and, for example, what parts of London do the French visit the most? When people from Spain visit London, what are the parts that they tend to visit? You can see strong segmentation. In the field of urban planning now, the use of geolocated social media is becoming one of the biggest new datasets that is being used to understand the flow of people.

And if you think about this, it raises the notion of a fusion center—a globe that shows everything that is happening around the world, where things are happening, and the information behind it. Can we actually get to a point like this? Is it conceivable that we could build something like this?
A lot of my research focuses around this concept of can we leverage machines to process what is happening around the world and then construct catalogs of what happened, who was involved, and how they felt about it? Imagine a system that can take a news article about Iraqi leaders who criticized Turkey on Monday for air strikes against Kurdish militants in northern Iraq. The machine reads this sentence and reports that there are two things going on here: One is that the Iraqi government criticized Turkey, and two is that Turkey bombed Iraqi Kurdish rebels.

This is very powerful. When you think about that sentence, a human can read it, but even a human takes a few seconds to read that sentence and figure out what it is saying. Now imagine not looking at a sentence, but at the whole news article. Now imagine multiplying that by every news article published in the entire planet each day. That is an awful lot of news media. No set of humans could possibly read through that much information.

The other nice thing is when you move beyond English. Think about the number of languages that are used. Humans have to be trained to understand languages. Machines can build a version of an algorithm on any language and, all of a sudden, you can force leverage that data. Essentially we get a spreadsheet of what is mentioned in this article. From that spreadsheet, we can make a map and represent what is happening in Egypt right now. The spreadsheet can also map the unrest of last year. When we produced this map of Egypt, it circulated widely. I was getting phone calls and emails from people within the U.S. government, impressed and interested in what the map presented.

Some people say, “Look, we have our own feel for what may or may not be happening there. What we want to ask is that you find all the news media that is out there and just give me a map. Just give me a map of the consensus view of all the world’s medias telling me what is happening there.”

One of the key elements that I am hearing the most in terms of this big data revolution is its ability to predict anything and everything. Like a weather report, “Riot, five o’clock next Friday, on this street
corner.” The problem is when you think about the weather, you do not call it weather prediction. You call it weather forecasting. The weather person may get up on TV and say it is going to pour rain tomorrow. And tomorrow comes and it is a beautiful, sunny day. But the point is that they are right more often than they are wrong. And certain weather forecasters are better than others; in certain areas of the country, they do better than others. The important thing is that it is accurate enough that we can actually take some action on it.

You also note that weather people rarely give forecasts twenty years out. They rarely say the weather here in Georgetown in the year 2050 is going to be 65 degrees and sunny. They do very short-term forecasts.

Can we predict the world? When I talk to people about this, it is very interesting because that is what everyone wants. At the same time, there is a general consensus emerging that while forecasting is great and certainly can have a lot of power behind it, what machines can definitely do today is give us real-time situational awareness maps. The latter is an ability to be able to peer into things.

When you think about a country like Egypt, it would be great to say what will happen there three years from now? Or, with all the mass executions that were just ordered, are they actually going to carry those out? Those would be great things that the magic 8 Ball could just tell us.

Certainly machines today can start giving us feedback. For example, during the peak of the unrest machines can put a dot at every location where news media around the world report on that unrest. That by itself is incredibly powerful.

Where this becomes exceptionally interesting are these two maps that I am going to show you. The first is a map of Ukraine. I did this map the Friday before the president fled the country. Every red dot is a location of conflict, of active conflict. Pink dots are locations of protest. And what is interesting is the size of the intensity of events there.

What was also interesting was the commentary I heard: “Well, you know, this map is not going to be
really interesting because there is nothing more that is going to happen there.” Feedback on the red dots in Crimea from very senior people was, “Well, this must be a mistake because there is nothing happening in Crimea. This is just some protest in Kiev. This really is not involving the rest of the country.”

That was enormously telling to me. This was not some random person saying this. These were fairly senior people saying, “A. Ukraine’s done; nothing’s going to happen there; and B. Crimea has nothing to do with what is happening in Kiev.” People said that must be an error. In retrospect, the president fleeing the following day was quite interesting for a map.

Looking at a map like this is not going to tell you that the president will flee tomorrow and that Crimea is going to be annexed by Russia. What this does tell you, though, is that this is more than Kiev. This is not some small isolated event in Kiev.

When you start drilling into some of the media coverage points near Kiev, it is very interesting, especially these two points. What you start to see is a local radio broadcast that said, “There is going to be reduced police and fire service for the next few days. A convoy of buses took all of our available police and fire services to deal with the ongoing situation in Kiev.” Then additional local news media reporting the same thing. That is a very powerful indicator that the government is going to try and do something here.

Of course, what happened was the president fled. But again it is a very powerful way of being able to sort of see what is happening across the country—a machine’s eye view, but of media. That is what I think is most interesting of where we are today. We absolutely interpreted our natural world.

We have satellites everywhere. You think they must be observing the entire earth. Just turn back the clock and you can see anything that is happening anywhere. What we are finding is that even with satellites, it is not a perfect world. We can only observe bits and pieces of our world.

When it comes to the natural world, there are great databases of earthquakes. Any earthquake that occurs anywhere on the planet today, you know
within minutes. As we have been seeing with all these aftershocks in Los Angeles and now in Chile, we know pretty rapidly what is happening where. But when it comes to society, if I say give me a list of all the riots that took place in the world yesterday, we do not have those type of datasets. In the human world, we do not have that catalog as well.

Here is a similar map; it is the last one, made from Nigeria for the past year. You will notice a lot of red up here, in the area where Boko Haram is, and not a lot of protest activity. A Boko Haram arms depot was destroyed; you can see its location on the map. To be able to see these conflicts, to watch what is happening in the country, is a powerful capability.

The question of course is how closely does this map match reality? One of the most interesting things for me is when I send maps like this off to various field desks and other groups and ask this question, the answer I usually get is, “Well, we do not have a map of an actual view.”

One NGO I did a pilot with in Afghanistan said, “Look, we’ve got field offices throughout Afghanistan. We get reports from all these field offices, but we don’t have a consensus view of what’s happening there.” Another analyst emailed me and said, “my group is responsible for this particular country. We knew that there was some unrest in the neighboring country but our job is to focus on the country that we are assigned to. And so we’re watching the conflict here.”

So I had sent out a daily report each day, summarizing what had happened around the world, and included maps of movement. The response was, “Your daily map one morning actually gave us some critical information.” But then he said the desk officer assigned to the country next to us is watching their country. They are aware that we have some problems, but they are focusing on what is happening there. It was not dawning on them that the unrest was marching toward the border and that the part of the border it was marching toward was a very restive area. If the unrest reached this area, it was going to cause enormous problems to this country.
This is one of these powerful examples, especially in the intelligence community, that we are typically very siloed in responsibilities. People are assigned a country or an issue, but they are not noticing the fact that the two conflicts are about to collide with each other.

This Nigerian map is animated. If I move this time slider back and forth, we can watch what has occurred over the last three months. So this Google earth file is basically anything that has been recorded from the world’s news media that occurred in Nigeria since January 1, 2014. Items that have a green halo are cooperative acts, as in aid or agreements. Red depicts conflict. Things that are black are of a more even mixture. Basically every day, it counts how many things occurred in each location by day. What this is telling us is the daily ebb and flow, the pulse of a nation. That is a really powerful concept to be able to peer into the pulse of a nation.

I tried the same activity with the twentieth anniversary of genocide in Rwanda. Yesterday I started receiving tons of emails from media outlets asking, “Can you give us some visualizations on Rwanda?” So I created this little visualization here. Here is everything from the ‘80s to present. And if we look all the way back we can see a period of relative quiet. Then we start to see things popping up. Then the intensity grows. And then we start to see much more intensity.

As you can see, Rwanda was not one tiny little area, one little set of attacks in one little area. You can see how much of the country was drawn into conflict. That has real policy implications. If you are trying to understand, for example, Ukraine and what is happening in the capital, you may see that the rest of the country has nothing to do with this. That is much easier to intervene in, to try to constrain and to stabilize.

This is again where I see big data as having enormous possibilities. For example, this is a daily map that I produced that looks at everything happening around the world in the last forty-eight hours and then comparing that to the previous forty-eight hours. It is a daily heat map of countries showing
more or less stability.

So India, for example, lit up red hot with all the sexual violence issues coming to the forefront. People were saying, “Enough is enough. We need to deal with this.” So the ability to really look at it, not just what is physically happening around the world, but how are people reacting to it, can be very powerful.

Think about Syria. The government may actually be winning now in Syria. How is the world reacting to that? Does the world say, you know, “This is great. The Assad regime deserves to win. They need to crush the rebellion and destroy it.” Or is the world saying, “Assad really needs to get out of there. That government no longer has a right to rule.”

CONTEXT

It is interesting that we can start looking at the contexts of events. This slide is a new visualization that we built, in a very compact way, to look at change over time. The ‘x’ axis represents days. From January 1 to December 31, each row is a year. And it goes from 1979 to present. And at each location it is sizing a dot based on the intensity of what was happening there.

This is looking at the unrest in Egypt from 1979 to present. Here we have Sadat’s assassination. We can see his assassination: the burst of unhappiness, then the pause, the assassination. We can see the terrorist turmoil here, with a lot of the bursts of what happened there. But we can see the 2011 revolution, in
terms of the global history of Egypt and the media reports, what happened in Egypt was a lot less intense than what happened at the assassination.

You can see, of course, the two bursts from last year. And you can see that, yes, there is stuff happening in Egypt right now. But compared to the revolution, compared to last year, compared to what happened here, you can see—based on how the media is portraying it—that it is not as intense.

This is a fascinating graph. This is by day, going upwards. This is every country in the world, minus the United States. You are looking at the intensity of a discussion around sovereignty. But what is interesting is right here. You can see Ukraine and Russia, in lockstep, with a high intensity discussion of sovereignty.

Both Russia and China have long, constant, ongoing discussions of sovereignty, especially in China and Asia. Can China be a sovereign nation? Will Taiwan and China respect each other as a sovereign nation?

So the use of machines gives us the ability to come up with new ways of visualizing data. You imagine taking a year’s worth of global news media, taking every country in the world, and saying, by day, what is the intensity of discussion in that country with respect to sovereignty.

Now, this is another slide I did for the Washington Post, a sequence on word clouds last Monday. I took each world leader and sought the top one hundred names that have occurred in news coverage of or about that world leader.

And what you get, for example, is Toomas Hendrik Ilves, president of Estonia, obviously a strong connection to the Baltic States. You can see Dalia Grybauskaite, president of Lithuania. What is really interesting here is where is Paul Krugman? He is a *New York Times* columnist who got into a very famous Twitter spat with the president. This is Putin’s word cloud. You can see the former Ukrainian president there. You can see a strong connection to Edward Snowden. You can see a lot of very interesting sort of contextual pieces of these leaders.

These are two interesting ones. Chavez’s legacy
with Maduro is very strong here. What’s quite interesting here with the Rwandan president is Patrick Karegeya—I am probably butchering that pronunciation—who is actually a Rwanda dissident who turned up dead in South Africa. You can see some very interesting contextualization—how the world’s news media contextualize a given leader.

NETWORKS

And that brings us to networks. You know, when you think about a leader you think of how do they exist on the world stage? This is actually Edward Snowden. So this was an interesting question. It was WikiLeaks that helped him flee to Russia; they bought him his plane ticket and really helped him. And one of the stated goals at the time, one of the things that WikiLeaks did at the time was say Snowden is going to be our new brand. He is the new face of WikiLeaks.

But then Snowden was elected number one on Foreign Policy’s global thinkers in December. And what’s really interesting about that was WikiLeaks reacted very negatively to this. They did a full court press, saying, “Snowden did not deserve to be in that, it should have been Assange, not Snowden. Snowden hasn’t really done anything of merit.”

And that was very interesting to me because the organization that helped him escape, that put him on the podium, is now attacking him. What does this mean? Why is this?

So I started digging into it. This is Snowden here, this is obviously Obama. What is very interesting is you will notice Snowden took all the news coverage over the past year on government leaking, spying, and NSA [National Security Agency]. It turns out that Snowden really has become dead center.

You can see this purplish-blue. That is Russia. If you notice Russia is not dead center. When people talk about spying, Russia is not dead center. Putin has been very good about managing media coverage, “Russia’s giving him sanctuary, but we have nothing
to do with him. He’s just a political dissident here.” And so that is a very powerful part. But what is most interesting is down here. We can see WikiLeaks, Assange, and Manning being pushed outward. And then you can see some of the other leakers, like Drake and some of the others, pushed further down.

What is interesting up here in blue—these are all the prominent journalists that are covering this space—you would see the journalists detaching and reattaching on Snowden here. From a policy perspective you can see the news media is really recentering on Snowden. Snowden’s brand has replaced WikiLeaks’ brand.

Looking at one particular case, let us say that you are the State Department trying to do messaging on topic. What you want to know is, how is the dialogue centering? Is this a WikiLeaks thing or is Snowden really separate? What this is really telling you is WikiLeaks actually is dying as a brand. WikiLeaks is no longer getting coverage. When people talk about leaking today it is about Snowden. WikiLeaks rarely gets mentioned in that coverage. So that is a very interesting dynamic.

The ability to look at Iran is very interesting. What you see are the outward layers, the people that interface with the world who are most closely connected to people like Obama, Netanyahu, and other leaders. What you see is actually successive connections. People that are, say, in the military hierarchy or successfully inward, will actually be successfully inward in that network.

If you start drilling into this network, it starts matching the political hierarchy of the nation. And what is very interesting here is that the Supreme Leader is largely absent from this network. By choice he stays out of the media spotlight.

This is a different analysis on social media and we fine very fascinating things. When the United States think about launches in Iran, it is about nuclear ambitions. What is interesting internally is that nuclear ambitions really are not being discussed much. Internally, the subject is political dissidents, torture, and arrest.

So this is a very fascinating ability, for the first
time, to be able to reach inside a country and compare what people inside the country care about? Think about this from a policy perspective. If you are the United States and you’re trying to encourage talks with the Iranian people and say, “Here’s what Iran’s nuclear ambitions are,” well, you are not going to get an audience there because that is not something that is cared about as much.

Here’s another example. This is the Nigerian oil and gas industry, and looking at each person in the industry: basically taking all the coverage over the last year of the Nigerian oil and gas industry, including mention of every person in the industry, and how they are interconnected.

It turns out you can start getting a very interesting structure. You see the heads of industry and certain governors that have heavy interests there. Through calculations you can show who is the most influential. It turns out that a couple of business people are more powerful than the president. And, of course, that often times is the case when you come to oil and gas and energy.

This is a network diagram I did in December for Foreign Policy. It was a year’s worth of global news media, putting every person that was mentioned in the world’s news media into one massive network diagram. You start to see these fascinating communities and clusters of people. That is very powerful, because it allows us to better understand society.

EMOTION

Emotions are particularly critical. This is a very interesting graph, on “Obamacare.” If you scoop up all the American news media coverage of Obamacare
by location and color code it from bright red, which is very negative, to blue-green, which is positive, you could start seeing real negativity spread in July of 2013. Then it starts reshaping toward more positively. By the time of the government shutdown, the whole country is turning very, very negative.

You might say, why do you need a super computer to tell you this? What is interesting about this is that you can start to put numbers behind anecdotes. At the time I put this out, people walking around D.C. were saying, “Well, it seems like there is a lot of coverage of Obamacare.” It is a different animal when we are able to say, “Well, 81 percent of American television news programming has at least one feature on Obamacare right now.” That is a very different animal. That puts us into a world where we have to respond to this. Here is how it is being contextualized. Here is how it is being connected.

Another interesting feature of this was that a lot of Republicans viewed it as a defeat. In reality, it was very successful for them in that the media coverage had really died off about Obamacare. By linking the government shutdown to that, you find a handoff. When the healthcare website issues happened, the media took that as a handoff. The Republicans had essentially turned the media coverage to say, “Cover this as a negative issue.” The media obliged, seen clearly through these numbers.

This is a collaboration I did with the Syfy network. Where they said, “Look, we want to do something that’s never been done before. We want to measure social media coverage.” Gnip was the vendor that powered this. What we did was to take the live feed of all the tweets about the show and then design a new way of measuring the emotion of those tweets.

We took every character on the show and measured the emotion around those characters. We had a very, very complex formula behind this, but it
measured all the emotion, how the emotions were trending, all these different factors, and came up with a popularity score for each character. This information was updated every ten seconds. You could follow this on your iPad during the show. It was really great.

In the opening show they had to go up on the platform with cattle prods. They had body armor to protect the heart so you could not kill the other person. You work on a platform, cattle prodding each other. Again, I had nothing to do with the contents of the show. During one of the other episodes they tied someone outside in the yard, overnight, without any clothes on, to freeze them as punishment. Now, it was a very interesting show. But what was very interesting about this was that very first episode.

There were two people facing off. One is a decorated Marine, injured in combat, a war hero. The other person is just an average Joe. And this average Joe was very popular. Everyone loved him. Well, the first show they are getting up to go on the podium and Joe just jumps over and shoves the Marine off the platform. The Marine falls twenty-five feet and breaks both legs. He is out before he could even start. Within seconds, without seeing this on the show myself—I was watching all the technical details—the lead character was not J.R.; it was one of the others go du, du, du, du, du, du-du-dudududu. His emotional stats start flying down to the bottom. Then he starts really diving downward. That told me that something major just happened.

But then, the very next night—they’re twice each week—people said, “Wow, you know, maybe he was just being overly competitive.” So then you saw people start rising “Joe” back up again. As I had been showing this around town, the number one request, especially from embassies, was, “Well, could you do that for world leaders? Could you show me how my world leader, my head of state, ranks?”

So the world leaders are next. We took all the world leaders each day, and rank them. I include the head of state and the next highest official, but the one right below as well, so you can see sometimes
the segmentation that happens there. You will notice most of the ranks are negative. That tends to be the case. What was interesting is you can see some clear breaks here. You can see some clear segmentation.

This was during some of the Thailand demonstrations, so you can see he is way out there. What is interesting, though, is that people said, “Well, it’s not necessarily how negative a leader is even if the whole population is against the leader. If there is no credible opposition nothing is going to change.”

The next version will measure all of the political opposition. We will create a new ranking that will give you not just the overall tone, but how the tone shifts.

I did graph I on Assad. What is interesting is this part right here, Aleppo. Now, right here Assad is in free fall. This is what is known as the free fall of death. When a leader enters a free fall like that usually they do not exit except by exiting the country.

But then you notice the numbers are more positive. Now, again, he is still in the negative, but he has just massively recovered from his previous position. Especially when you look at all the news coverage during this time period. When Aleppo happens and the whole world says, “He’s done.” “He’s going to be nuked to nonexistence.” Or, not “nuked,” but completely removed. But then nothing happens.

About 72 hours later is about the magic mark when, all of a sudden, you start seeing across the world news coverage saying, well, “If he can do that —” No one provides even a symbolic statement. No one sends a cruise missile over his head to make a statement, even though he is vulnerable at this point. He is all-powerful now. Because now he can just gas all the rebels and nothing will happen. That was very fascinating because you really see this in the data.

Now again, for Ukraine this started off with incidents of police brutality. It led to some interesting questions around the country and, of course, where the situation is today.

Putin is another great example. So you can see Putin kind of ebbing and flowing. Post-Crimea he is gaining in global discussion. And this takes into account not just people saying, “He’s a great person”
but also “He’s invulnerable” and “He’s all-powerful.”

If you look at other dimensions on Putin, you see that globally his stature around the world is strengthening, that essentially Russia is invulnerable again.

One of the ways that we found this to be very useful is to be able to look, for example, if the U.S. is participating in a trade summit, the U.S. may know, “Here’s the reality of what’s actually happening. We’re the ones that are actually driving this trade summit.” But how is the world portraying this summit?

In one particular case, global media was basically saying that the United States was along for the ride and it was the French putting all the policies forward. The U.S. has lost its ability to command the discussion in this area.

That becomes a very useful dynamic if the State Department says, “People aren’t perceiving us as being the power figures here. Maybe we need to issue some press releases or increase our public statements on this dimension so that people will recognize that we still have a leadership role here.”

The Fortune 500 are trying to create graphs that map the tone about their products; they are doing things. That has not yet made its way over to the public policy space.

Countries are not doing that as much. A lot of that is just getting bureaucratic momentum. Some of it is fear. A year and a half ago I was giving a presentation here in Washington, D.C. and was asked, “What is Twitter?” I thought it was a joke. I actually laughed it off. But it was not a joke. You still have people that do not understand these things.

I have sat in many meetings in D.C. where people are using English language tweets to look at the Syrian rebels. Another example, people speak Russian in Russia. They tend to post in Russian on VKontakte; this a very important platform in Russia. When you are only using English language Twitter or English language Facebook, you are missing a big part of the discussion.

In particular, the biggest problem to me today really is that, on the one hand, some companies
say, “Big data is whatever we sell.” And on the other hand, you have people saying, “Big data is evil. Quantitative research is evil. Everything should be qualitative. There should be no quantitative, no big data. Computers are evil.”

The trick is that you need a blend of ideas. You really need that blending because in a chart like this you can find an infinite number of patterns in things.

Like that Nigerian map, the data can say, “Well, according to global news coverage, here’s what’s happening in Nigeria.” What you need to do then is ask if this is reality? You need to work with people in the field to see if this matches reality, what are the ways in which it does not match, and why might that be? And the other piece then is let us say you this matches reality. Well, how do we start translating that data?

This is the ultimate question with dashboards. Anthony Alcott and I have a National Academies grant looking at dashboards for intelligence and policy. Dashboards are designed along this notion of a closed system. Think about an airplane cockpit. You know if you turn the knob this way, this thruster is going to fire, and the plane will go in this direction. The laws of physics hold.

In countries, there is no law of physics that applies. You do not say, when I do this action this will result.

Oftentimes it’s unclear what to do. Take Ukraine, for example. Even if you had the perfect system that predicted in two months’ time that Russia is going

---

2. “In management information systems, a dashboard is ‘an easy to read, often single page, real-time user interface, showing a graphical presentation of the current status (snapshot) and historical trends of an organization’s key performance indicators to enable instantaneous and informed decisions to be made at a glance.’”

“In real-world terms, “dashboard” is another name for “progress report” or “report.” Often, the “dashboard” is displayed on a web page that is linked to a database which allows the report to be constantly updated.”

to take Crimea, what do you do? You are likely not
going to send troops. You are not going to risk nu-
clear war over Crimea. How do you deal with this in-
formation? So that’s an interesting question by itself,
even if we had the perfect system.

Of course you also end up with things like para-
dox of prediction. If you have the perfect system that
tells you exactly what’s going to happen when and
it’s accurate enough, then basically what’s going to
happen is if a system says there’s going to be some-
ting happening here, you take action to prevent
that. Well, the machine’s going to learn that under
these circumstances this type of stuff doesn’t hap-
pen. So you end up with all these issues.

But really, all those are theoretic at this point be-
cause we are in the first stage of use. We have all this
amazing data, we see the corporate world do all this
great stuff, but from a policy standpoint we are still
dipping our toes into this world.

Today, we are starting to explore what is possible.
We start by having demonstrations of how we can
empower citizens and how we can understand the
world in new ways.

These are, I believe, all the exciting areas of where
big data is going to take us.
Panel Discussion

Comment and Question: I teach a class in communications, culture, and technology and it was interesting to learn that a number of my Middle Eastern students are on Twitter, but would never think of using anything like their real name or real location.

I know that DARPA [Defense Advanced Research Projects Agency] did this incredible red balloon project about three years ago. It was a project to see if the crowd could find these ten red balloons that they put around the country in public places. And the most fascinating thing they found was that a whole bunch of people got on in the Twitter feed and started posting totally bogus information about where the red balloons were. And hundreds of people were off chasing balloons that did not exist in various corners of the country.

How do you sort out that? If people are going to start using this for mission-critical decision-making, there is going to be even more disinformation out there. It is not just people protecting their privacy. It is nation-states intentionally trying to deceive. It is going to be pranksters. It just takes a couple of bad pieces of data that get repeated because they are so interesting to really skew some of the results that we are showing.

So, any ideas on algorithms for truth?

Dickover: Two points on that. How many of you follow Andy Carvin? He is this awesome person for curating news, and he has all these contacts in Libya that he is communicating with, with Skype and so forth. And when something hot and heavy is happening I will usually open up his feed as a separate line because he will do, you know, 50 posts in a 20 to 30 minute period. The stuff he is talking about will show up on BBC like four or five hours later.

But talking to the point about real identities. In many cases, he is actually retweeting people’s real names. Most of the people that signed the Declaration of Independence knew the risk they were taking. Right? And we remember John Hancock. Some of these folks made the case, “I am risking my life for
this and I want to be recognized for doing so.”

So again, it is back to this notion of a calculated risk. There are some that do put their real names forward, primarily because they know that they are going to be associated with it and they are not willing to take the risk.

In relation to sort of the dark side of use of social media, you have seen the debate between Evgeny Morozov and the other side. It is the new battleground, if you will, where authoritarian regimes are trying to provide noise on the channels, and you have the other side, who are trying to think through circumvention technologies and other ways to make it happen.

The real problem is with the non-technical folks. If you are a peacebuilder and you learned three months ago at a conference or some gathering that, “If I use Tor in this way for my contacts, I am safe.” That does not mean it is going to be the case four months later. Because the information environment is changing so rapidly, this is really problematic for folks who are not technologists, to think through a change that could put them at risk. That is just a very difficult problem.

**Mayo:** From my perspective, one of the things a lot of our customers are interested in doing is identifying what is spam and what is not. So what ends up happening is that by filtering out what could be deemed as spam, they are missing a lot of the communication that is actually occurred in the distribution of the information.

So there is a paradox that even a lot of our customers struggle with to figure out. I think the *Wall Street Journal* had an article last week about Twitter, where they reported that 40 percent of new accounts in a day are created by bots.¹ It is a challenge for us. It is one that we have chosen to not really tackle with a technical solution as of right now, because we still see that there is value in understanding that component of things. Even though it may not be indicative of a human that is there, in the end humans are driving that and making the information spread.

So it may not answer your original question, but it is something that definitely factors into our business.

**Lectaru:** You definitely see this. One of the ways that Ushahidi’s been reacting to this is to get a trusted account. If something gets posted it does not get the gold star of a trusted account unless it is someone they have worked with or unless it is also showing up in the news media. Patrick Meier, who was head of outreach for Ushahidi, now head of Social QCRI, is focusing on new algorithms to try and detect this. What they are finding is that false information tends to spread through Twitter in different ways.

For example, look not just at a piece of information, but the way that information spread. What are the characteristics of this message—who is sharing it, how are they sharing it, and how is it entering that network. That is how people are starting to look at spam on Twitter, astroturfing,² which is basically when you are posting all this fake stuff. Well, usually what happens is a big cluster of bots are talking to each other. There may be hundreds of thousands of people talking to each other, but basically you can kind of see how the messages resonate and how it is shaking outward.

One of the greatest ways you deal with the truthfulness of a message is by looking at multiple sources. For example, the Associated Press Twitter account was hacked. The White House account has been attacked. And the automated bots of Wall Street started exploding. What was interesting about this is that not all of the Wall Street systems started going crazy. A number of them basically have been built to look

---


at multiple sources. So you have this trade-off.

When you have something breaking on Twitter— I mean, Twitter often knows something before anyone else does—a better way to proceed is to hold off just a fraction of a moment and see if anyone else talking about this. For example, one of the firms I talk with said they have algorithms looking at the fact reported to them. Yes, AP reported this but, you know what? AP is not the only press office in the White House. There are a lot of press outlets in the White House. Not one of them is reporting this.

A researcher from the Library of Congress looked at how best to just propagate items across multiple platforms. They looked at Twitter, Google-Plus, Facebook, and Wikipedia. Most social media sites do not have an administrator. There is no one on Twitter correcting errors. Wikipedia has that. What the Library of Congress found was at times when something starts trending on Twitter, and it propagates to Wikipedia, the Wikipedia administrators start correcting the item. This is a pretty good indicator, pretty quickly, that there may be a problem with this source.

Another clue to the truthfulness of an item can be found if it starts spiraling across multiple languages and no one is rolling it back. As we see it across Facebook, Google-Plus, and others, our confidence in that information goes up.

That, to me, is common sense. The bottom line is that we have never had media that moves information this fast. Today the goal is how do I get it faster and faster and faster.

That is something that even affects us with machine processing. Look at the Boston Marathon: the bombing that occurred in the United States on the finish line faced an enormous amount of media coverage. Within the first fifteen minutes you had reports for anything from none dead to one, I think, reported almost a thousand dead. The problem is that because all of a sudden there is chaos, smoke, and blood, and everyone wants to be the first one out there. Even the news media is no longer saying, “I am going to wait.”

Question: You mentioned Evgeny and I just wanted to say that he is also a former Yahoo Fellow at Georgetown. My question for the panel concerns the sources of information that you use. This information comes from companies who are responding to consumer needs. They are going where the customers are. You either take feeds from the companies or follow the business demands for the products. What is your sense of the responsibility of companies to protect human rights in that context or even to promote them?

Dickover: The responsibility they have is to protect the connection to their platform.

Going back to Burma, I was talking to this transparency organization, a small organization. The guy that heads it up left the country after the Saffron Revolution. He is working with USAID [U.S. Agency for International Development] to do documentaries of how this change has occurred and, you know, this is sort of a rumor society. He said one of the big problems the activists had here is that the government had access to Russian hackers. They were able to get into our content really easily; it is a very normal thing for somebody’s Facebook page to get hacked.

He told me that this happened both to him and, more importantly, one of the biggest youth leaders in the country who had over 80,000 followers—there are about 500,000 people on Facebook in the country. His account gets hacked; he stops using the page; and creates a new one. Now he has 5,000 followers. This is a completely normal response. The guy that I am talking to did the same thing.

You start asking questions, “Well, how are you accessing Facebook?” They have never even heard of SSL encryption. If the government owns the ISP and the telecom, they are seeing your user name and password in clear text. It is not really that hard; they
do not need Russian expertise to hack your platform, even though HTTPS3 works there.

Edward Snowden said 80 percent of the deal is to encrypt your communications. So I think the hope is not just in the U.S., but other places, that things like that encryption are done by default so somebody does not have to make a personal privacy security choice to access that.

But just an anecdote about the consumer and electronics. Depending on how you measure in Burma, there is somewhere between two to five consumer electronics or consumer IT companies. If you buy a Smartphone you are buying it from a cell phone store. And they are going to charge you an additional 30 dollars for them to load everything on your phone you want. This is all the music, all the apps, everything, but really they are going to set you up with a Gmail account that you never use, and to get you a Facebook account. That cell phone is your primary platform.

Astrology is very big there. If I hear your name I know immediately what day you were born on. That is embedded in the name. So this guy puts up this astrology app. The way he does it is to send these tokens to the cell phone stores. When they install it on the cell phone, it is an additional $2 charge for each one of these apps. He gets the money directly because they do not have banking.

Technology is not driving a lot of these truthfulness concerns and, in some cases, there are even questions hate speech online. Is it occurring after the fact or is it driving the conflict? This is a hard question. When you get into some of these places it really does change dramatically.

From my perspective, it is just doing the basics to take care of things like that for consumers. I do not think there is any way they can police the whole content, but at the same time we want to work with folks like Facebook to think through how we can address inflammatory speech in places like Burma prior to an election. Facebook is fairly open to that kind of stuff.

Mayo: This is a really, really critical issue for a lot of the publishers of the data. Consider Twitter—I will take them because it has probably the largest dataset. In order to be able to distribute data from our platform to a customer, Twitter has the right of first refusal to serve any of those customers. There is a lot of scrutiny put into new customers that are going to potentially buy data from us, to understand their use case, the potential application of it.

And they, I believe, have put a pretty strong focus on understanding user expectations and what data should be distributed. For example, we do not provide any private data. The only data that we are providing is public data that you could see from a user’s posts on Twitter. If they have enabled geolocation within their account, we will provide that as well, though we do not provide IP addresses.

Gnip has a contract with our users when they sign up for our platform that we can provide their data to a third party. It is very important that we honor that contract and think about what a user would expect with that.

Because for Twitter—and Facebook is in the same boat—their number one metric for their investors is growth of their platform. And if their users start to feel concerned about their privacy being violated and the fact that Twitter is using them and particularly putting them in any danger, I think that they will flee from the platform. People will begin to disband it. So it is a really big issue for these companies to think about.

The revenue from data quite honestly is not a large component of their overall business model. I think it is a significant one, but I think they are taking a conservative approach to how the data are distributed and want to ensure that user rights are respected.

Leetaru: Ultimately the question boils down to who makes that decision about what is hate speech. You look at Syria today. Syria has declared that any anti-regime elements are terrorists attacking the na-

---

tion. Anti-regime speech is terroristic speech.

Look at Crimea. How many of you have seen those pictures that Russia put up all across Crimea showing the Nazi flag: “This is what happens if you go to the EU. That’s the Nazis. You are going to be part of the Nazi party.” And then the Russian tricolor is shown in a beautiful park. And Russia itself has been banning a lot of anti-Russian material. The first thing when the Russians troops arrived was to shut down all the independent media centers across Crimea.

The United States cannot make these decisions speech or news within the country. Look at China or Turkey right now, which have said that even the discussion of government corruption threatens national security and is terroristic speech. For a nation to say this has to be banned because it is threatening our national security, how do you draw a distinction and say that it is not really an okay ban for you to do?

Al-Qaeda is making heavy use of social media for recruiting and for messaging. This presents an interesting question: Do you try to shut that down or do you let it continue? Historically these groups have been meeting in cafes where we cannot hear what they are saying. Today they are doing it online where we can at least observe them, see who comments, and the messaging that they use.

What it comes down to is censorship. It is a big discussion on the Internet, concerning freedom, and DNS. There is the view to allow everything and let people decide. It faces a competing view in countries that are saying any messaging posted by anyone around the world that is negative to my regime is illegal and I want it removed from the site.

This presents an interesting situation. What if social media companies start censoring material? What happens if Egypt says, “Any anti-Egyptian material is considered illegal.” What if Facebook or Twitter agree to remove any negative comments about Egypt? You run into a slippery slope. Where does the censoring stop?

From my perspective, social media companies should ensure a connection that governments cannot block and to ensure the privacy of people that are on there. Give the people a voice; let them make that choice.

**Question:** Starting with Noel, you were mentioning the technologists that you work with in different countries. I was intrigued by that. Who are they? Are there commonalities among them? And my broader question is, what would you like to see in terms of training and education for people who work in this fascinating new world?

**Dickover:** I will compare technologists, which provides a commonality across the world versus peacebuilders. The work of peacebuilders is to establish a long-term, enduring relationship. There is an expectation that this will be a multiyear process to build trust and then to start moving the different parties together.

USIP has a series of Justice and Security Dialogues in an Iraq, Libya or Burma, or wherever else. They will have a group working from the civil society perspective, maybe some working with the police, other groups in government. At some point in the future, they start to build ties between those two or more perspectives. The idea is if we can build trusted connections, when tensions arise they will have an avenue to deescalate that. And that is a long-term process.

Technologists really meet and build trust as a series of one-night stands. We get together during conferences and get an immediate connection with somebody. For certain parts of my life, I am going to trust them implicitly and immediately about some

---

4. “The Domain Name System (DNS) is a hierarchical distributed naming system for computers, services, or any resource connected to the Internet or a private network. It associates various information with domain names assigned to each of the participating entities. Most prominently, it translates easily memorized domain names to the numerical IP addresses needed for the purpose of locating computer services and devices worldwide. The Domain Name System is an essential component of the functionality of the Internet.” Source: Wikipedia, http://en.wikipedia.org/wiki/Domain_Name_System, April 23, 2014.
things. If I am not in Arabic countries, we are going to have drinking sessions and we are going to connect. Then I may not see them again until an event three months later and we connect right again.

The way you do trust-building in the technology world is very different. It is like I quickly recognize this person as “of the tribe,” and there are a lot of things that we can correlate on. That is true in Honduras, in Dakar, or in Jakarta. It is a sort of affinity; people are going to talk about their achievements as a way of gathering stature.

It is really weird if you go to a meeting of peacebuilders where there are multiple groups that have come together. There is this really interesting sort of ritual, if you want, where whoever holds the meeting has to give a ten to fifteen minute soliloquy that shows detailed knowledge and deep understanding of the circumstances. Then a member from each other group will pipe up and add additional content, where they all say I know what I am talking about.

When the peace tech folks are meeting with groups that do not understand technology, they do not bother with that level of detail. But the way that they build trust and affinity, and the way the peace-builders do it, is a very different process. The good part about it is, again, you can build a small level of trust very quickly and really start working toward that. But to have the tech for social good piece and to have them connect with development folks is a longer term process.

**Mayo:** I will save time for Kalev to talk about it because he is probably more involved with some of the connections to these folks.

But from my perspective it is really being able to leverage open source code, repositories, and sharing best practices for interacting with different modes of data. For us, our developer, Evangelist, will spend a lot of time working with folks to really be able to build better interfaces to our systems and how to integrate it into other platforms. We have just put a lot of code out to get hubbed so that folks can use that across the globe.

I think that is probably the most tangible way that I have seen it happen for our APIs and our data in particular. I think we will continue to see that grow as more that people interact with us.

**Leetaru:** One of the big things is data literacy and map making. Fifteen years ago the only people really making maps were geographers. They were the people that had GIS [Ground Information Systems] training and knew how to think about all these issues. The rise of things like Google Maps and Keyhole have totally transformed the world. Today people think in terms of maps. People walking down the street have maps, live maps, on their phones with satellite imagery.

This spatial literacy has changed so much. For example, look at a discipline like history. One of my close friends and mentors is a historian; his dissertation was the first digital Census work on the old punch card machines. So he has kind of seen that whole field develop. Twenty years ago, what we saw in the history of general maps was usually battlefield things. It was troop movements. Today you are much more likely to see maps of just about anything, because the literacy is there. People know how to think in terms of maps.

Data is still an interesting space where you think about how to use data today; the tools are not really there. The problem is there is no magic, data toolkit. If you are doing network analysis, you are using either igraph or Gephi. If you are doing heavy GIS stuff, you are probably using ArcGIS.

If you are doing statistical analysis that is low-end—that is statistically intensive but not computationally intensive—you are probably using R. If you are doing high-end computational intensive analysis you might be using MatLab or Mathematica with the grid-enabled capability. If you are doing really intensive stuff, you are probably writing your own BLAST algorithms.

The problem is that it is still so fragmented. There is the technical side of how do I program this analysis. Then there is the side of actually thinking about data itself, about asking critical questions. When someone says, "Hey, I made this great map of what the Syrian rebels are saying
about the government . . . ,” do you know enough to ask, “Well, what data is this based on?” If the response is, “Well, this is all English language Twitter,” you have to be able to say, “Why do we believe that this represents the Syrian rebels?” And then when the person says, “Well, because it has the word Syria in it . . . .” The big problem today is you have people that know the discipline, but do not understand the data.

One of the greatest dangers today is that the data is so beautiful; it is so pretty. Someone gives you an Excel 1.0 graph. You are going to look at it and say, “I have some questions about that data.” But when someone comes back with this beautiful, Macintosh Apple-style visualization, where the numbers jump and fly through the air, with maps twirling through the air, it gives you a false confidence. Like, this visualization looks so beautiful, this data must be real; this data must be perfect. If you called ten people on their cell phones and ask them how they feel about the government, the problem is you can make a heat map from that that makes that look incredible. The difference is also interesting to me. People at the State Department want their memos; they want text; the want one paragraph.

And then you talk to the analysts, they say, “I do not want any of the summary text. I want the maps. I want something that is a static map.” Then you get all the way down to a line analyst who says, “I want to dig into this stuff. I want to be able to drill; I do not want a map in my hands. I want the original data, to be able to look at it myself.”

When I think how do we build interfaces, how do we think about all the different pieces? Who is our audience? What goes into that? What are the datasets that go into this? These are all these things that we are having to still explore.

Open source software is something that is really novel. Someone comes up with some really cool algorithm and they can share it. Someone can say, “Hey, you know, I built this great data algorithm that takes Gnip data and does this, this, and this, and is open and free. So anyone can come along and say, “Hey, I am using this data. Hey, here is this great tool that does that.”

If you want to make a heat map today, you are not having to say, “Boy, I have to learn KDE algorithms, interpolation, and quadratic code.” Now you grab this toolkit, feed in an Excel spreadsheet, and you get a heat map. “And geographers have vetted and said that this algorithm is valid.”

Question: What could researchers do to make their work more machine readable?

Dickover: In our terms it is a combination of this rich, deep discussion with datasets, and the process of releasing this. The list of people who really want to take advantage of machine-readable information in developing countries is not really vast. They are not, “Wow, a new dataset. That is just what I wanted.” There is a neat example in Kenya. One of the countries that really focuses on the government open data movement. There was a massive flurry of activity once they first released all this stuff. The World Bank and others got to do it. But it falls down fairly quickly when people realize the government is not really responding to this. As an example, what if I create a system where I identify all the potholes. If anybody has been to Chisinau, Moldova, every street would have a big red block on it, because nobody has done anything since the Soviet Union left. The whole street is potholes.

To the notion that the government is going to magically respond to citizen engagement, that is a hard process. Like in Iraq, we are looking to follow up on these peace tech exchanges, to do a series that connects provincial government folks to technologists just so they can start thinking through the process of accountability, transparency, and citizen engagement. We would look on how that would benefit them. Hopefully, later, do the same series with civil society and then connect all the parties.

For the first event, we brought some government folks to these more inclusive meetings. It was complete oil and water. Literally, in the inter-
acting online session on violence against journalists, the government said, “No, no, let us just have an open conversation.” The civil society folks said, “Well, of course they would do that. What did you expect?” It just fell right into stereotypes. Even though we had gone through this four-month vetting process of picking the people that the embassy and others said were trusted agents. There are a lot of issues well before you get to machine-readable stage.

The part that we want to do is engage in this conversation and it’s back to the local data. That is what they have that is critically valuable. If you could start aggregating, making sense of, and connecting to some of these datasets, we believe it starts getting down to ways that you could help them with tactical decisions.

If these things are not helping somebody think through the scenarios and leading to behavioral changes and expectations of future changes, what is it really doing? I do not buy that Obama is going to solve problems in Country X. It is going to be the participants of that country. How do you enable them?

For us, it is this conversation with rich text and now informed by data. If curators and people can teach them how to use this on the ground—these peace tech roles—they can help train people in data proficiency, why it matters, and ask them questions that hopefully provide an understanding of information. I think is the short answer; it is really difficult.
About the Speakers

NOEL DICKOVER  
Senior Program Officer  
U.S. Institute of Peace

Noel Dickover is the senior program officer who heads up the Open Situation Room Exchange project at the U.S. Institute of Peace as part of the broader PeaceTech Initiative. As an internationally-experienced senior consultant at the intersection of technology and development, Noel has consulted with a number of Federal agencies, including the State Department, USAID, Administrative Conference of the U.S., Health and Human Services, and the Department of Defense.

For the Department of State, Noel initiated, developed and headed up the eDiplomacy Division’s Civil Society 2.0 TechCamp program—TechCamps are international, highly interactive two-day events which partner technologists with civil society organizations to raise their digital literacy with low-cost, easy-to-implement tools to improve their mission. Noel has brought the TechCamp approach to USIP (now termed PeaceTech Exchanges) in a modified form which now include back-end funding.

As a co-founder for CrisisCommons, a global network of technology volunteers working on disaster relief efforts, Noel helped lead a movement of technology volunteers that has transformed the way regular citizens can impact crisis-response efforts in a disaster situations, starting with Haiti. He initiated a number of Government 2.0 efforts for the Department of Defense, including authoring their draft social media policy which allowed soldiers to use social media such as Facebook and YouTube.

KALEV LEETARU  
Yahoo! Fellow in Residence  
Georgetown University

Kalev H. Leetaru comes to the Institute for the Study of Diplomacy (ISD) as a “Big Data” and high performance computing scholar with the support of the Yahoo! Fund on International Values, Communications Technology and the Global Internet.

Before joining Georgetown University, Kalev held the Irwin, Boyd Rayward, Josie Houchens, and University Fellowships at the University of Illinois Graduate School of Library and Information Science, where he was also Assistant Director for Text and Digital Media Analytics and Senior Research Scientist for Content Analysis at the Institute for Computing in the Humanities, Arts, and Social Science and Center Affiliate of the National Center for Supercomputing Applications.

He holds 3 U.S. patents (cited by 44 others) and his work has appeared in the press of more than 100 countries. His most recent work includes the first in-depth study of the geography of social media and the changing role of distance and location in online communicative behavior around the world (named the top social media study of 2013) and the creation of the GDELT Project, a database of more than a quarter-billion georeferenced global events from 1979 to the present on the people, organizations, locations, and themes connecting the world. Most recently he was named as one of Foreign Policy Magazine’s Top 100 Global Thinkers of 2013.

WILL MAYO  
Managing Director, DC  
Gnip

Will Mayo is the Managing Director of the Washington, DC office for Gnip. In this role, Will is responsible for managing a large group of commercial and public sector customers in the Washington Metropolitan Area, as well as identifying new opportunities for organizations to leverage social media data. He is passionate about data and analytics and believes there is limitless potential for improving business decisions with the use of social data. Prior to joining Gnip, Will was based in Colorado and served as a Senior Manager in Deloitte Consulting’s Strategy and Operations group. In this capacity, Will provided strategic planning, customer relationship management, and data analytics expertise to address various client needs.
Georgetown University

Founded in 1789, Georgetown University is a distinctive educational institution—a national university rooted in the Jesuit tradition of social justice and education of the whole person, committed to spiritual inquiry, engaged in the public sphere, and invigorated by cultural pluralism. Georgetown’s location in Washington, D.C. provides a unique platform for Georgetown faculty to make their expertise and talents available both to policy institutes in Washington as well as to a wider international audience. No other American university is better positioned to foster a critical dialogue on global issues.

http://www.georgetown.edu

Edmund A. Walsh
School of Foreign Service

Georgetown University and the School of Foreign Service exist in the most fertile international arena in the world, allowing the School to establish globally renowned competitive programs and centers as well as offer first class undergraduate and graduate degrees. Founded in 1919, the School remains committed to educating students and preparing them for leadership roles in international affairs.

http://sfs.georgetown.edu

Institute for the Study of Diplomacy

The Institute for the Study of Diplomacy (ISD), founded in 1978, is the School’s primary window on the craft of diplomacy. The Institute’s constituencies include diplomats, scholars, and Georgetown students. ISD staff and associates teach courses, organize lectures and discussions, mentor students, and participate in university life. The Institute also convenes conferences and working groups, and sponsors and publishes research. ISD international affairs case studies are used in classrooms across the United States and around the world.

http://isd.georgetown.edu

Master of Science in Foreign Service

Georgetown University’s School of Foreign Service conferred the first graduate degree in international affairs in 1922, pre-dating the U.S. State Department’s adoption of the term “foreign service.” Since that initial class, over 3,000 students have completed the Master of Science in Foreign Service (MSFS) Program. Graduates have attained notable success in careers with national governments, international organizations, private businesses and civil society groups.

Today MSFS is recognized as one of the most selective programs in the world. As reported in the January 25, 2012 issue of Foreign Policy, a recent survey of international relations faculty ranked Georgetown among the field’s top professionally oriented master’s degrees.

http://msfs.georgetown.edu
Information on ISD:
http://isd.georgetown.edu
Institute for the Study of Diplomacy
Edmund A. Walsh School of Foreign Service
Georgetown University
1316 36th Street, N.W.
Washington, D.C. 20007
Telephone: (202) 965-5735
Facsimile: (202) 965-5652

Information on MSFS:
http://msfs.georgetown.edu
Master of Science in Foreign Service Program
Edmund A. Walsh School of Foreign Service
Georgetown University
37th & O Streets, N.W.
Washington, D.C. 20057
Telephone: (202) 687-5763
Facsimile: (202) 687-5116
The online world is allowing faster, more representative, and round-the-clock access to societal behavior around the globe. Today, people from Bangladesh to Buenos Aires busily tell one another and their neighbors what they see, what they think, and what is important to them, offering unparalleled visibility in the digital heartbeat of global society. The proliferation of always-connected mobile devices means citizens and participants are often the first to report on emerging events, streaming photographs, videos, and ground reports as events unfold, while social platforms have become one of the primary organizing tools for rebel and opposition movements. Moreover, the constant stream of daily life that flows across online media platforms provides rich contextual background information on the narratives of each region and culture.

A growing chorus of voices argues that “big data” is a usable, practical tool to help foreign policy makers, implementers and analysts increase their understanding of global issues. For example, computerized analysis of massive amounts of world press reporting yielded an innovative map of Ukraine’s protest violence useful to any individual grappling with the present and anticipating the future of that region. Can similar analysis of the “thoughts, discussions, and debates expressed in public social conversation” provide similar insights for foreign policy practitioners, scholars and students?