ABU DHABI – Open Data Initiative
Thursday, November 2, 2017 – 10:30 to 12:00 GST
ICANN60 | Abu Dhabi, United Arab Emirates

ALAIN DURAND: Good morning. Welcome to this Open Data Initiative Pilot session. As you can all see, I'm not Ed Lewis. Ed is unfortunately unable to make it today, so he asked me to fill in for him. I'm going to do my best to impersonate him, but I cannot claim I will do a good job. There are a number of times where I hope I will not say something too stupid, and if you have questions specifically on some of the wording he has on his deck, I will try to answer, ask some of the other panelists to answer for him, and if we don't have an answer, then we will get back to Ed and bring the answer back to the community.

So, next slide please. Yeah, let's go to the next one. Here we go. So, the first part of this discussion today, I'm going to present this slide deck that Ed has created for us. Then, we will have a short panel with Suzanne Woolf and Marc Blanchet who are working with Ed on this and they will explain where they are in this process. The most interesting part is going to be the community discussion. So, we'll have four panelists that I will invite, and then we will spend the rest of the time going through their presentation and hopefully we'll have a lively discussion at the end of it.

Okay, so next slide. So, open data. How many of you here are not familiar with open data? One? Okay, everybody knows about Open Data or at least has heard about it. So, open data is a pilot at this moment. It's something that is run by the office of the CTO in ICANN and this is in order to figure out what the full open data initiative is going to be. So, we are trying to understand what we need to put in there. The objective of this pilot is not to provide all of the data immediately. That's not the expectation, but to understand how we're going to be able to sustain this over time.

We'll talk about some of the lessons learned, and at some point we'll have to hand this off to the appropriate team within ICANN. That essentially means the IT department within ICANN who is going to run with this project. So, hopefully it will not simply be us sending this thing over the wall, but there will be some kind of a transition phase and there will be some discussion to avoid some surprises, and there may be a bunch of activities that will be coming after that, so next slide please.

What is open data? The whole idea is to make available to the community a large number of data and data sets that ICANN has that can be made publicly available, and I will go into more details about that later, what does it mean that can be made available. Today there is a number of data already available of

our reports, there are a bunch of web pages, but this is not always sufficient. Why? The first thing is those reports are not machine readable. That's perhaps one of the main things here, is we want this to be ingestible by machines so that people can make their own statistics. We can take something from a report and the report next month and the report next month and the report the month after, and see what the trend is. So, that's one thing. If you simply have a PDF, that's doable, but it's kind of difficult sometimes because it involves a manual process. We would like to avoid that.

Timely delivery. Well, it's not guaranteed at this point. So some things are going clockwise, some are coming whenever they can. There's no automated process right now to really deliver this. Some of the data is not complete in scope and some of the data we have is simply not published and people have been asking us for this data so we have to figure out what part of it can be made available. For example, when we publish some reports, there's data that is used to build that report. How much of this can we make available, and how much of it does the community want us to make available. I have to find a trade off here.

Next slide. If you look at the current data that is available today, it's, in Ed's word, spotty. It means that it's not the same everywhere. Some projects have more data than others, and in

some cases, it's easier to retrieve than others. So, one of the goals of open data is to provide some kind of a common interface for all vessels so it will be much easier to retrieve it. There is a lot of data that is scattered across web pages or PDFs or PowerPoint, not the same format everywhere and doesn't make it easy to retrieve.

As I said, some data is downloadable, some is not. Some projects have their own open data initiatives, and that creates duplication of efforts that may or may not be the best use of resources. Some projects have data available upon request, so if you know who to ask, you can get the data. If you don't know who to ask, you can't. It may not be the best, again.

Next slide. What's wrong? My point that I've just made is, you have to know who to ask to get the data, so if you're part of an inside club, then you have access to it. If you're not, it may be more difficult. We'd like to fix that. So, incomplete data is a problem because you try to derive conclusions for something that is not a complete set, and you may miss things. Scrapping data from a PDF file or a webpage is what we did last century. In the 21st century, it should be all about APIs and other ways to automatically have a machine-readable data. Once again, this is to go from a manual process to an automated process.

Next slide. So what's the focus of this effort? Well, automation, I think I've said that a number of a times already, try to make this automatic, make this machine readable, downloadable. When I talk to Ed, he says, "Computer readable. That's the one thing I want to derive from here. Computer readable." Interfaces to this and the other thing he insisted on was the repeatability of a process, as in there may be at the beginning maybe a small number of data sets, but there will be more, and we would like to have a process for which we decide how we are going to take something that is proprietary and make it available, what are the ways to analyze this, to look does it conform with privacy laws and others, and how do we look at this? How are we going to store it? How are we going to share it? It always has to be a pipeline process, and if it's the same for all the different data sets, we'll be in a much better shape.

So, vendor selection. I will talk about this some more, but at the end of this project, if we have a good idea of what could be the proper vendor to make to put this in place, that will be mission accomplished.

Next. How much data do we have at ICANN? The answer to this is we don't know. That's not a good answer. The first thing is to figure out what do we have. Traditionally, nobody has really tracked how much data we have. We have data to do our job.

Data was a tool, not a product. So, we are somewhat changing that perspective. When we have data now, we have to look at things like GDPR, the data protection regulatory survey. So, that's another effort that is happening, there is some IT stuff happening and looking at all this data, so Suzanne will talk a little bit more about what we have in this census now.

Next slide, please. When we started this project, people were saying, "Oh, you have like pots of data available. Just make it available to everybody, not just internally. That should be easy, right?" A pile of gold there, just distribute it, be done. Well, not so. Here's the problem. When some of the reports are made, sometimes they are made from private data, data that we bought somewhere. Sometimes it is data that is encumbered with privacy information [inaudible]. We just cannot take this and make it available. It just doesn't work. It's not that easy. Sometimes data is in a database.

And next slide. Here it is. That's the Human Resource database. It's one large database. Gendered diversity. It is in the report, but this is from information that is in the database that we cannot necessarily show. So that's one of the complexities here. There are also things like salaries. Some of the salaries are reported, some are not. So, there are laws that regulate that, we cannot simply take the data and make it all available. It just

doesn't work this way. It has to be a process that is thought about.

Next slide. Now we come to an interesting question. I've been talking about data set for a while, and what's a data set? Again, if you look at [inaudible], sometimes the data [inaudible] that's been used to create this report is not a single element. It is the result of several layers of processes. We start with one [inaudible] is going prepared, processed. It goes to another machine that processes again, but merges it with data coming from another place, and all of this in the end produces the report. So, what's the data set there? Is it the final report? Is it the different layers? We are not necessarily sure.

So, let me give you an example. The project that I run called ITHI for Identifier Technology Health Indicators is looking at a number of metrics about DNS and about other things. In the end, we may have hundreds of metrics. Is the data set the entire collection of all those metrics? Or is the data set one particular metric down to tree and we have like hundreds of data sets?

So if you say, "Data set is what I want to track," you can say, "Well, data set is the last metric because that's what I care about." Or you can say, "I want to have a holistic vision and data set is the entire system." So, it's not obvious. Now, when we talk about vendors and vendor negotiations, the price that we

pay vendors is usually charged by data set. So if it's like 1,000 data sets versus one data set, it's a completely different story. So, that's something we need to take into consideration here.

Next slide, please. So, can everything be a data set to be managed? Well, it's not obvious again; some data sets are required, like I was talking about the mandatory reports. So, we have that. Some are in development. For example, I have some other research project that I'm running and we have a bunch of data for this and we're trying to understand what is behind all of this, some research questions. Should we publish all this data? Maybe, maybe not. My personal take is if we have some findings, it will good to publish some of the data that led to that finding if the data is not encumbered by contracts or by PII. Some data sets are temporary. They're something that is being arranged for five minutes and then they're gone, so is there any value in publishing those things? Maybe not.

Next slide. So, I talked about this a little bit earlier with some data sets are encumbered by contracts, and some are encumbered with privacy information. Those cannot be published, so I talked about this already.

Next slide. The first part of this is to identify all the data that we have. There is a census that is being done by Suzanne, and identify what can be made public, and then prioritize it, because

if we end up with a large number of data sets, we may not want to publish it all at once. We may want to start small, and then increase from there and try to make it better because we are in the learning phase. If we do go with one, we may learn that that didn't work exactly the right way, we have to do it slightly different, the second time we do it better, and hopefully by the fourth time, we will know how to do it right.

The idea behind this is once we get this IoN out, we will have a process that any new data set will have to go through in order to be published, and then hopefully the process will be agreed upon by the community, we all understand what it is and there are no surprises. That's where we would like to be.

Next. The census, again, Suzanne will talk about this, so I'm not going to say much about it.

Next. The pipeline, Marc will speak about this later too, so I'm not going to say much about this.

Next. The prototyping activities. There was an early pilot developed in June. We had four vendors that were contacted and participated to this pilot, and the goal was to get a feel for what those tools were able to do. It's great to say we have a tool, but a tool doesn't magically solve a problem. We have to

understand what are the limitation of these things, design philosophies and all of that.

The current thinking is that the tools for open data are not specific for ICANN. All kinds of industries are using similar tools to publish data that are held internally. We believe that those tools are mature enough so that now it's about selecting one that fits for our purpose and we should be okay. We don't have to necessarily develop an entire thing.

Next please. This is the discussion. The current assumption is we will be able to find one of those tools that fits the purpose and work out a process with IT to integrate it into the rest of the tool chain.

Next slide, please. So, when we have those different vendors and different tools, part of the project now is to evaluate them, and to evaluate those tools we need to ask questions. We can ask questions from a community perspective, what does the community want from the tool? We can ask the question from an internal perspective, what do we internally need from the tool in order to do the job? So those are sets of questions, and I would encourage if we have time for discussion for you to think about what are the things you would like from the tool.

So, questions are, I need to ask Ed trying to think about what the community would want, so it's me trying to think what Ed wanted to say about this. Can data sets be downloaded in a bulk? I want all the data, not just the data from January 17, 2016. I want the entire set for the entire year, and I just want for this particular thing, but I want it for the entire set of metrics that are related to it. Can we have this bulk transfer of the data?

Can data be interpreted or am I simply going to get from a tool a bunch of numbers and I have no clue what those numbers mean, or is there a way in the tool to actually graph this data and very quickly see if there is a trend or not? Or can I take this set of data, that set of data, map them together and see if they correlate? And how much do I want that because it can be as fancy as you want.

Sometimes those correlations are extraordinary complex and better done offline, and sometimes they are simple and can be done there, so it's a trade off of the more we put into the tool, the more complex the tool becomes, the more expensive it is, and the more brittle it is. The less we put in there, the less useful it is. So, where is the right balance here?

Can we dig into the data? So, again, let's do the reverse example. I started from, okay this is my entire set of the year,

can I go and see what happened on January 26th? And can I save any of that?

Next slide, please. Flipping the story and looking at it from an organization perspective, can we present a narrative around the data? Can we tell a story? Can we say, "Okay this is the data we've been looking at. We from ICANN believe this data is showing this, this and this." Can we put this in there, and can we provide you the tools to actually look at this and say, "Yeah, I think this guy is right," or, "No, this is not correct, and we want to send you some feedback if there is a problem."

Can we do complex things, like saying, "Okay. This data, this data, and this data; I have three layers of them, and I want to show visualization that integrates all those three as what I will present to the community." As an ICANN owner of the data that I want to present to the community, can I do this complex integration? Can I use different media? Maybe I have an HTML5 file that has a dynamic graph where you see bubbles going up and down. It's what I've been personally working on recently. Can I mix that with some CSV file and mix that with a basic Excel graph?

Next slide. So, this is a URL for the early pilot, and those are the different vendors that we have been using. So, one of the pilot was actually an internal tool, and the other vendors were

Enigma, OpenData Soft, and Socrata. So, it was the first one that we used in June.

Next slide. In the early pilot, there were a couple of data sets that were chosen. For the sake of simplicity, we took some registry monthly reports activity and registry monthly report transaction. We just had to start somewhere. That's what we picked.

Next, so there's a mailing list if you want to provide feedback. That's the URL for this. The slides are available on the web, so you don't have to take a picture or write it down, you can go and download it. There was predictably very little activity on the mailing list because there was not much publicity made about this at the time. Unlimited data to play with, so now we'll expect that we can get more activity. It was essentially a starting point for this discussion.

So, next slide. We are entering our second round of pilot. That is due by the end of this calendar year. We would like to put more data set, and focus not so much on the graphics and things like that, but focus on the [inaudible], focus on the pipeline. Are we going to make sure that the data gets from where it is to where it has to go? How to make sure that this gets automated. The two vendors will be part of a second round of pilot. One will be OpenData Soft, and the other one will be Socrata.

Next slide. This is still a research activity driven by the office of the CTO. This is not yet production. I mentioned that at the introduction of this talk. At some point, we will have to hand this over to the IT department so that they can run this into production.

Next. There are some milestones coming up. There's an internal report that's supposed to happen this week for whatever definition of this week is, meaning that people are traveling so it may be pushed by one week. That's an internal census, an internal report. The second pilot, December. Around March, we think that we will have an internal census and an internal pipeline; all these will be internal documents.

There might be more pilots that we want to run starting March or April next year, and finally, the goal is to understand this enough to involve the IT people at the end of June, 2018. At the end of the fiscal year 2018 so that started fiscal year 2019, IT can run this. The details of this hand off are not yet very clear. I think a lot of it will depend on the outcome of the pilots and what type of tools will be selected.

Next. That's the last one. Thank you. So, that's the end of me pretending to be Ed. Ed would have certainly been more in depth and more forward about this. Thank you all for indulging me trying to be him. Next, I would like to invite Suzanne Woolf

and Marc Blanchet to talk about where they are in the census and the data pilots. So, if you'd like to come up. So Suzanne, I will hand the microphone over to you. So maybe introduce yourself and then --

SUZANNE WOOLF: Yeah, because I've never presented to an ICANN meeting before. I just have one really substantive slide if you want to go ahead and -- I'm sorry? I have it. How cool is that. So, I'm Suzanne Woolf. I've been associated with ICANN in various capacities for as long as there's been ICANN. Currently what I'm doing is I'm actually performing the data census for OCTO, for the Open Data Pilot, and Ed asked me to sort of add some color to what was in the previous presentation.

Alain actually covered most of the really important stuff about the data census, and I'm really eager to get to our panel, so I just wanted to add a couple of high level points. The census is actually a really key step to making anything happen with open data inside the organization because you have to know what you have before you can decide what to share. So, the data census is an organization internal product to enable further management of the data. What it's there to do is basically to get from where we are to where we have to be in order to decide what data can be released, and currently, the data census was

undertaken because there's been no unified tracking of data sets across the organization and there's been no unified procedure for documenting what's considered public or not.

The thing I want to emphasize here is that there's nothing unusual at all about this. With a medium sized organization that's extremely complicated and has many constituencies, so many moving parts, it's not at all unusual that this work has to be done before it makes any sense at all to talk about open data. The mechanism that we're using is a survey just provided to all the departments and just asking for the metadata about the data that people work with, both the technical description (where does the data live? In what formats? How large? How frequently it changes?), and the administrative description (Who uses it? Is there PII included? Is the data set currently public? What reports are based on it?).

Because a big part of the purpose for doing this is so that when reports are produced, when the organization shares information with the community, that we be able to point to where that information came from, how that report was generated just to improve the clarity and transparency of the work product that comes out of the organization to be able to show their work.

The next steps, just for where we're at, literally, the working draft of the census lands any day now, but the next step beyond

that is revision. It really is a working draft, because we're finding out new things about questions we should be asking and so on as we go. I think that's why it's octo and not production yet.

In addition, the other major next step, and I think Alain touched on this briefly, is an internal procedure for vetting what to release, and people feel very strongly that the default should be to transparency, obviosuly given the nature of the organization, but there are also reasons to hold back for much the same as, for instance, with the documents release policy. So, the next step here is a procedure for deciding what can be released and making sure that it's in the accessible format that we've been talking about. And Marc can talk a little bit about how to, once data is identified, how we are making it available. [AUDIO BREAK]

MARC BLANCHET: Good morning. So, we're doing what was referred in the presentation by Alain as the second phase of pilot or something like that, and the idea here is to make a prototype of the actual pipeline. So a pretty short presentation to give you an idea and one of the reasons is we just started, so the overall goal is to prototype the pipeline from data sources to output for open data providers.

Some requirements is kind of an easy, fast, dirty prototype, meaning that we may not [inaudible] production level kind of stuff at the moment because that's going to be a prototype, and the idea, again, was to learn enough so that it will be good input for the real production environment that will be taken care by the IT department inside ICANN. And that the output itself would be provider agnostic, so there is currently two targeted open data providers, but we want to make sure that the whole pipeline could be used by any of them or all of them, so we want to have a standard common interface to them, so we're open data provider agnostic.

We will be working with a few data sets from different sources, hopefully with a variety of data types so we exercise more all the issues that we can encounter. Some possible normalization of data. We'll see how it goes. Some possible data-processing. We'll see how it goes. From basic stuff such as date formats or [inaudible] or anything. We don't know what will happen, but we'll see as we go. Early results as soon as possible, we just started the project a week ago, and we have a short deadline of December.

This is the generic architecture. You will have a few data sources. Some may require some adapter. We want to have this stage here kind of more generic. The adapter here could be

either something here or more near the data source to be discussed. Some data processing, as I said, normalization data or something. Staging the database and then a common interface, as I was talking about, that would be vendor agnostic, so ICANN would be able to use whoever open data provider they want.

Current work. So, defining interfaces, the messaging, the formats between the different pieces. Key design consideration examples are for, are we doing pulling of the data or pushing the data from the sources to the pipeline, as well as through the open data vendors. Is this going to be "real time" or "almost real time" or "scheduled" or "triggered?" It might depend on the data source. This all has to be identified.

Identifying who is doing the work on the data, preparing the data. Is this the data source or the pipeline or both? Some normalization, messaging. Stating on the various technologies. Pick your right technology. Do you want Python or Java or whatever. Database, QL no SQL. Identify the data source and data sets and doing the actual implementation. And I think that's my last slide.

ALAIN DURAND: Thank you very much, Marc. I would like to maybe hold off questions toward the end, and at this point invite the four panelists, so if you guys would like to move up front, and we will start the panel. [AUDIO BREAK]

So, not in the order in the panel, we have Jay Daley from NZRS. We have Roland LaPlante from Afilias, Christa Taylor from DotTBA, and Jonathan Zuck from CCTRT. So, the first presentation on the panel will be from Jay. So Jay, why don't you introduce yourself and take it from there?

JAY DALEY: Hello. I'm Jay Daley from the DotNZ registry. We have been running an internet data portal, an open data portal, for the last three years. That's the URL you can see for it. The background product is Socrata, which is one of the products being looked at. Now, I'm not going to go into the details of why I open data is so important, but I assume most of you here know, but if not, please ask during the questions.

Ours is slightly more than just domain name related. It's intended to be a portal for all data about the internet in New Zealand that we can find. It is all open data. Some of it is ours. Some of it is taken from third-parties. We split that into four categories: Domain names - which is our data, Internet

connectivity - so that's everything having to do with ISPs and the type of connections and the number of connections and the technology involved; then global rankings - so that's anything that we can find that ranks New Zealand compared to other countries for internet usage or other things, such as the web index or the cloud readiness index. Then, Users and use is where we put all of the results of our surveys that we conduct and that other people conduct on users, how they use the internet, their perceptions of the internet and other things. This is really quite intended to be a comprehensive data set on here.

The dotNZ data sets, there are four of them. The oldest one is the dotNZ Zone Scan, so monthly we scan the 700,000 domains in our zone and we look for 50 or more indicators. We then upload aggregated statistics for each of those scans. So, no domain name is individually identified, but for example, the number that have a particular type of broken delegation will be listed there as an indicator. That's useful then for people to track the overall health of the zone, and trends in how the health of the zone changes over time.

Then we have the DotNZ registration data which is updated daily. That's the number of creates we had yesterday, the number of cancellations, the number of particular types of transactions all the way through. And again, that's aggregated,

but that allows people to have an understanding of how our business is going without, again, seeing the details of each domain name.

We also do a regular scan of SSL and TLS certificates in dotNZ and we put up aggregated stats about that, including who the certificate authorities are so that you can see the rapid growth of Let's Encrypt from our data, for example. Then, all of our name servers are instrumented and do full packet capture and we pull that data back to a large headed cluster. Recently, we finally added a publication of open data related to that so the number of particular types of queries we see and those sorts of thing.

This is taken from our Zone Scan, and this is just a simple chart available on the site that shows TTLs and basically just showing, I don't remember what the different colors are, but there are peaks of TTLs below five minutes, and then a big peak at five minutes and then another peak at 10 minutes. But there are some people that have some bits in between. This is just a little snapshot. So, that's the type of interesting data that you see.

That's useful when you're trying to calculate what happens if you completely destroy your zone, how long before the internet in your country ceases to work. That's what the underlying data looks like, so we have the classification. Those are MXTTLs and

the metric shows the bracketing off from there or the billing of them and the number of domains in each one and the count.

So, the lessons that we learned doing this for a number of time is that the API axis is vital, and that's both for us to upload the data to it and for other people to download the data. The reliability and speed of that API matches tremendously. One of the biggest issues we've had is the speed of the API for the amount of data that we want to upload to it, and ensuring that we can do that and do that well.

There are other things in the product that matter to us. The nature of the catalog. Simple things such as whether it has persistent URLs and what they look at and the data formats and the way you can get the data out. There are lots of different things to consider there.

The other thing is that we look at long versus wide data, and we process everything to be long data, and actually, we denormalize rather than normalize as much as possible because we want the data to be available for people to take a single data set and use that. Having someone use multiple data sets to get a result is just complex and so we really try to avoid that. If we duplicate, that's fine, we don't mind. We're not building a database, this is really quite separate to that. So, we have to do quite a bit of pre-processing.

Now, just to go back a bit, this is a long data set. There's a different way that this could be represented that would have multiple columns. Quite often when people create data sets, they think about multiple columns and they put things in there, but that's not the way data scientists use it, or good data talks process it. You reduce the number of columns and have greater numbers of rows in the way you do it. That's vital, again for people to be able to extract the most from it. If you put too many columns in, then you're telling people how they should use the data, rather than giving them the opportunity to use it as they wish.

The biggest problem we have is changing source data. That's potentially either within our organization or within third parties. Every time our local government statistics agency publishes a new data set relating to ISPs, I have a little cry in the toilet, and then I employ a student who comes in and does the messy work of turning that back into the data that we expect.

Finally, the biggest thing was the cultural change of remembering to eat our own dog food. We have to remember, internally, that all new data sets go on to the portal. It's not an afterthought. It has to be a beginning thought. And if we need data internally, then we have to use the portal, so if anybody makes the mistake of asking where's this data, then six or seven

people tell them immediately, "It's on the portal, don't you know that?" And so, we've finally built that culture for that as well.

So, that's it from me. And that's a marvelous Adobe addition to my slide there. Who's next, Alain?

ALAIN DURAND: Christa, you're next. Jay, thank you very much for your presentation.

CHRISTA TAYLOR: Good afternoon. My name is Christa Taylor. My company is dotTBA. I do new applicants and launch their new gTLDs. It's usually all around obviously numbers, financial numbers, and obviously also about data. So, I do predictive analytics and wasn't sure on everyone's background on predictive analytics and all the rest of it, so I'm going to take a little bit of a step back and not go into it too deep.

So, just to kind of give you an overview. 54% of professionals surveyed felt that their companies would be more competitive if they used analytics to make decisions. Data helps you make an impact and utilizing it will only bring you forward. It helps you make better decisions rather than hunches. It will help us

choose or make our future new gTLDs and all that data a lot more successful.

It impacts our existing technology as well as the future technology, and we also have to be really considerate of the Data Protection Directive, the 95/46/EC, which is coming out in May 2018, so every company should be well aware of this at this point in time. I really believe that providing reliable, open, clean data sets are beneficial to everyone, and everyone should have that access to make their own decisions.

So as I was saying, I do predictive analytics and I'm sure you've all seen the movie Moneyball. It's a branch of analytics that's used to make predictions about unknown events. For instance, is your baseball team going to win. The different techniques include data mining, lots of statistics, modeling, machine learning, and artificial intelligence. As I was saying, instead of making assumptions and hunches, it allows us to be more proactive and anticipate the outcomes based on the data.

People who use it, for instance the marketers who use predictive analytics, have a revenue growth of about 2.9 times higher than the industry average, and 2.1 times commanded a leadership position across products and services. That statistic is by Forester Research.

There's two kinds of data. One is retrospective. It's already happened. So, it's based on the past. We have the characteristics and all the variables are there to be used. I'm sure we all recommend or automatically associate this with Amazon. When you go to order a book, obviously, they have a whole bunch of other books that we might be interested in. On the flip side, though, we also have predictive analytics, and so what happens if our customer changes? What happens if I'm looking at a book in analytics, what's the next step I'm going to go down to, and what happens if that changes, and what are the new trends, and taking that all into our data set to be able to use it and to guess what will happen in the future.

It also helps in market segmentation, so we can recommend who to target, and let's not forget, who not to target. We always focus on trying to get that customer into the door, but we sometimes go and alienate them on that second step and we scare them away. We don't really think about that when we're actually targeting our customers, and that's a huge cost.

It impacts our operational efficiencies, we can anticipate demand, our conversion rate, our customer relations, and obviously, our pricing strategies.

So just to give you a little bit of an idea here for business intelligence, we have our WHOIS, we have our DNS, we have our

renewals, we have our deletions, and all that data is currently there for us. On the right, we can see sometimes the variables in there that actually make a difference that we can figure out which ones we should incorporate, which ones actually matter, and which ones don't matter, and you have to actually go through the entire data set and do all of those correlations to actually figure that out, and to figure out the variants that you can actually account for.

So, I can go through a million lines or ten million lines of data and I can come up with five variables that will account for 25% of the variants. I will never cover the other 75%, and even if I bump up my five variables to 150, it might only bump it to 30% of the variation being covered. So, being able to, as we heard earlier, combine the data sets with other data sets becomes really valuable, and making sure that you have the entire data set is also really valuable. If you only have a portion of it, and you're missing those outliers, you're missing a great opportunity.

Here's some more utilization of it, and here we have customer intelligence. We can use it for marketing activities, as I mentioned before, on segmentation, and on the right here, you see sentiment analysis, so we can actually just run a program and say, "How is our company doing today? What does that

look like? Can ICANN use this?" For sure it can, because we can read all the data from say Twitter. We can see all that sentiment analysis and we don't have to do anything. We just pretty much push a button and we can see the pros and the cons or this particular graph is just on one article and how it flows between the positive and the negative.

Additionally, using it, we can obviously spot new trends, and I was mentioning before, we had to be really careful about the external environment, specifically around privacy. That's going to cause us a lot of grief in the future if we don't.

Overall, data is your company's new best friend if you want to make impactful smart decisions. Currently, less than half of a percent of all data is ever analyzed and used, and I'm sure ICANN we're full of data and we're not using it as we heard earlier.

Taking a step back, we have I think 571 new websites are created every minute. It allows us to compete smarter. And poor data can cost businesses a lot of money with 20% to 35% of their operating revenue. And retailers who leverage the power of data can certainly improve their operating margins by as much as 60%.

Two final remarks here, we can use it for future domain name technologies. We're all sitting here talking about the next round,

and in fact a couple of hours before here, we were trying to figure out how many new gTLDs we can implement every year, and we were just sitting there saying was, "We don't have that data." So, we need the data to make smarter decisions and to make more efficiency within ICANN.

Smaller companies with less than 1,000 people, 10% or more are likely to view data as a strategic differentiator. They're more likely to use it rather than their larger enterprises. So really, big data analytics isn't just for big businesses. It's actually for the small businesses to gain that competitive advantage. And I think we're saving questions to the very end, but I guess we'll see who is next.

ALAIN DURAND: So, that's the mystery and lottery of data. Left to right. Jonathan, you're next. I have to confess something. I did some random ordering this morning by numbering all of them in one, two, three, four, five, six, seven, and that's how it was ordered. I started with one and then Marc, and Susan with two and three respectively.

JONATHAN ZUCK: Good morning. Thanks for choosing our session over the GDPR session. We really appreciate it. The irony, obviously, is not lost

on us that we're here discussing open data while they're over there discussing closing it. I want to give the kindergarten version of Christa's presentation, basically, because I think the answer that we all have reached after her presentation, is you should hire her for your business. But absent that, eventually we won't be able to afford her, I just wanted to talk about things in a little bit simpler sense.

I run a non-profit called the Innovators Network, which is working on future of work issues and global unemployment, and different types of employment models around the world that might help to address that. But the reason I'm here today is that I'm also the chair of the CCT review which is the affirmation of commitments, now bylaws mandated review of the new gTLD program and its impact on Competition, Consumer Trust and Consumer Choice, and I will say that one of the biggest challenges of our review was that we decided to make ourselves be more quantitative both in our analyses and our recommendations for the community going forward, and getting data was a huge challenge.

It wasn't available, it was closely guarded, it was considered competitive, etcetera, and it was very difficult, so when we were looking at basic pricing data for example, and it in fact was a screen scraping exercise that we paid an analytics firm a lot of

money to go out and literally scrape data off of websites to get the pricing to understand the competitive impact of the new gTLD program. So, I speak as someone wounded by the lack of data that's made available and believe in its utility a great deal.

One of the words that gets thrown around a lot in this community and in other organizations is something called continuous improvement, and so the basic continuous improvement cycle is to identify the issue that you want to address, make a plan for how the current process can be improved, execute your plan, and then review how it went, and then start over again, right? So, if we think about the data driven version of this, it's not a build slide anymore. Okay. Sorry about that.

So the first word up there is quantify, right? So much happens here at ICANN where we say things like, "This program has been a disaster," or something like that, and then you actually go and look at the data and you realize that it wasn't the case. I wanted nothing more as part of the CCT review than to be able to point out some smoking gun and say, "This is what was wrong. If we just make this one change, then everything would be perfect," and we didn't find any of that. So, what I'm going to end up having to do with almost every face to face session, whether it's the IPC or the NCUC is tell them that all of their fears and

exaggerated beliefs are wrong. It's a lot of fun, let me just tell you. So, that's been our job.

But you quantify, you correlate, try to find what the correlated variables are for the thing that you're trying to address, because that's then the thing that you're going to try and influence through your plan, you execute a plan around that, and then you go back and measure the data element you used to define your problem in the first place to see if what you tried is actually working. And then, you reevaluate where you are, and either define a new correlation, continue as you are, but you just don't keep doing things because you decided to do them, if they're not actually working.

So, for example, from our team, one hypothesis that's come out and that's been said by a lot of people, spam has gotten out of hand with the introduction of new gTLDs. Well, when you look at the data, the level of spam is in fact the same, but what we learned is that spam is moving into the new gTLDs, so the overall level of it hasn't changed, which is one type of problem, but then it seems to be attracted to the new gTLDs for a number of reasons: lack of restrictions, high volume, purchases, name generation, low prices; these promotional pricing and things like that made it easier for spammers and so it migrated over to the new gTLDs.

So, once you identify your problem correctly, you have a better chance of trying to address it. So, we saw that spam is correlated to price, so maybe you're going to investigate something like create a price floor for volume of purchases, eliminate name generators at registrars. Who knows? These are the things you think about once you identify the issue might be priced and how you're going to handle that without overregulating the space. Those are some of the challenges that you try to face.

Then, once you've done those things, you measure it. Have spam rates decreased. That's why the data becomes particularly important, and if not, change course. So, that's the ultimate idea behind, in this case, more retrospective data in use in terms of continuous improvement, and it all requires data. So, we're big fans.

The other thing I wanted to talk about that folks have talked about a little bit generally, I don't know if this word gets used as much as it used to; is mashups, which is the ability to take data from multiple systems or multiple sources and merge them together so that I can look for correlations that might not be immediately apparent.

So, for example, back when Yahoo had something called Yahoo Pipes, you could take crime statistics and real estate statistics,

home prices, and be able to visually see the correlation that existed between crime statistics and home prices, and those came from two different sources. That kind of thing could be immensely valuable to us as well. So, Yahoo Pipes is no more, but there's all kinds of software like it that's fairly accessible to business users. So again, everybody that can't afford Christa can make use of some of these kinds of tools like Power VI or something like that from Microsoft.

Then you can begin to look for correlations. Domain sales and cellphone penetration. Domain sales and social media penetration. DNS Abuse and ecommerce penetration. There's all kinds of things that you can begin to look at that aren't just ICANN data with the data sets that exist out in the world with Bureau of Labor Statistics. What about unemployment and registration volume? It might be the inverse of what you think it is. It could be that people with no job are more likely to go register a domain name, right?

So, you begin to find those things when you begin to mix data together and that's why open data becomes so important. And I think that was it for me.

ALAIN DURAND: Thank you very much. Now we have -- no surprise our next speaker is Roland.

ROLAND LAPLANTE: I wanted to hear a Frenchman pronounce my name. It's awesome. Anyway, my name is Roland Laplante. I'm the chief marketing officer at Afilias. I've been at Afilias since it opened its doors in 2001. I'm just going to walk around a little bit. How many in the room have actually looked at this data and done anything with it? Alright, good. Because I was one of the original pilot guys, and what I wanted to show you is what I looked at. So, I'm looking at a small piece of a small piece of a small piece, but it's helping me to have a better assessment of the business and the industry that we're in.

So, just by way of introduction, Afilias is a registry service provider and a registry operator. In the registry operator space, we have a direct contract with ICANN to run the Info, Mobi, Pro and a bunch of other TLDs. As a registry service provider, we just do the technology, not the policy or the marketing or any of that stuff, or other ones like dotOrg or dotAsia, .arrow and so forth. We also support, do back end work as a TSP for 13 country codes.

We were quite heavily involved in the new TLDs and right now we're supporting about 210 TLDs with about 21 million names across them. Also, a bunch of global brands, and I'll show you why this is important in a minute.

If you look at the new TLD volume on new TLD stats, it got up to about 30 million and now it's back down to 23 million as a bunch of renewals did not come in. So it remains to be seen what's going to happen with the new TLDs and I know a lot of people are looking at this data.

But how about adoption and usage? Historically in this industry, we've said, "If you sell a bunch of names, you're going to be successful," but that's not the success metric necessarily that we're going to look at in all of the segments of new TLDs that we have. So, we've got SEO studies that the DNA has done that show that new TLDs don't hurt your SEO and may even in fact help. The Universal Acceptance Steering Group has just released a report in September that I commend to your reading that talks about the status of universal acceptance and how browsers are doing and how desktops are doing, and so forth in terms of recognizing both the new top level domains that are in and also the IDN versions of those, which is really important.

In an ODI, the Open Data Initiative at ICANN, I think is helping us understand some things about actual usage of these new TLDs,

not just how many we got sold. So, let me show you what I mean by that. I've only looked at the DNS data, specifically the UDP query data that comes in to our system at the name server level. These aren't all browser queries. These are just name server queries that we get, and we're required as a registry operator to provide this data to ICANN every single month. There's a whole host of indicators that we have to deliver every single month, and we've been doing that since 2001.

So all this data sits at ICANN and it used to be 20 gTLDs and we could pull the data from that pretty easily. But when it got to be 1200 gTLDs, it was impossible without an awful lot of work, and ICANN publishes this report on a three-month lag for confidentiality purposes which we could probably discuss separately. And for the open data initiative, the latest data, the most recent data is through February, 2017. It's historical data back to the start of the new TLDs and maybe a little bit earlier than that, but it ended in February, so some people have looked at the data and said, "Well, why isn't it more current?" and the reason is it's a pilot.

They wanted to see if anybody was going to look at it, if anybody was going to use it, if anybody was going to do anything with it, and that's what we're trying to do here. But this is the first cut at

the data, and I want to show you what the UDP data is showing me about what's going on in the industry.

So, this is actual data out of the ODI. If you look at the total gTLD-DNS query trend, you see in general it's up, which you would expect. The internet's expanding, you would expect to see total queries through gTLD name servers going up over time. We have this big spike at the end of 2016, and I'll get into that in a minute too. So, if you look at worldwide gTLD, and again, we don't have the data for ccTLD so this is just the g segment, you see that about 87% of all queries are in Com and Net.

Again, that's not necessarily a surprise, and all the other gTLDs combined are getting 13% of the queries, so Com and Net continue to dominate the industry in a pretty significant way. If you look at the trends, Com and Net is coming up over time. Other gTLDs, we had again the spike at the end of 2016, but you can see that Com and Net just dwarf everything else. Again, this isn't a surprise, but there's data that shows this.

What's interesting is, Com and Net percent of all DNS queries if you just look at the share, seems to have come down slightly over time, and what that means to me is that new TLDs are getting some adoption, and that the share of the DNS queries, the share of the queries going through the system and the

internet, is starting to shift to other things. I don't expect it to be a landslide, but it's nice to see a little bit of wiggle here.

So, let's look at categories. Then I looked at the categories of new TLDs to see what the trends were. In the generic TLDs, you can see that, and there's about 500 of these, they started earlier and it's much bigger than anything else. If you look at the dotBrands, dotBrands don't measure their success by how many registrations they have. They're looking at it in an entirely different way. So if you look at dotBrands, it's a lot less than the generics and there's almost as many dotBrands as generics, but you can see that's growing too.

There's a fewer of the geos, this is dotLondon, dotBerlin, and so forth, and are people adopting the geos? I sort of expected that the cities would have a faster adoption rate, that more people would be looking for city addresses just because they are so familiar, and it doesn't seem to be the case, but looking at the subcategories here, you can get a general sense of how much usage there is, how many people are querying this kinds of names.

If you look at dotTop all by itself, of course the slide here changes over time, but you can see that there's a dip here in the registration volume from nTLD stats and pretty much corresponding to that, we had a reduction in the number of DNS

queries coming in on dotTop. I just pulled a couple of examples here. XYZ, you can see it's very erratic here, and you can see XYZ's volume which was at 6 million and then dropped down to 2 million here recently, you can see that it got volume, that it was spikey, so I don't know what's going to happen with this, but you can see it's jumping all around. It's not very steady.

If you look at a closer look at brands versus geos with the geos on the bottom, I changed the scale here a little bit, you can see that the brands are outpacing geos in terms of traffic.

DotOffice, which is a Microsoft dotBrand TLD; it has five pages in it. It has a reasonably good trend here. DotGLE which is a Google dotBrand also has a reasonably good trend. DotRen which is a number three one has 328,000 pages in it, and you can see how that's picked up pretty quickly over time. DotITAU, which is another dotBrand, which has only basic WHOIS and NIC sites. You can see that the trend has come up and then it's been steady.

Monash, everybody uses Monash University as one of the examples of one of the more used new top level domains. You can see that trend. Barclays bank, you can see that trend. Again, up. They're all going up on the new TLDs. BNPPARIBAS, it was flat for a while, had a big spike, it came down but it came down at a higher level than previously.

DotAbbott, nice steady trend up in general, but -- and here's really the important part of this for me. There's some wacky data in here. This is not clean data. We're all required to give ICANN data. Every registry operator has to deliver it and we give it to ICANN for the ones that we're the registry operator for and we provide the data for all of our registry service provider customers as well, but let's take a look at some examples here.

If you look at the generics, I mentioned this spike at the end of 2016. If you look at the sum of all DNS/UDP queries across all gTLDs, you see this big spike here, but if you look at Uni registries 13 TLDs, you can see that the entire spike is related to Uniregistry. Now, I don't know, did they get DDOS'ed? Did they counter Yarc? I don't know what happened to this data, but I don't think that's real.

Let's look at dotBrands. So, the whole dotBrand thing, nice steady trend, very good, but what the heck happened here? Well, what happened here is BBC. You can look at just dotBBC, which was delegated in March, and that had a giant spike that threw off the entire trend for the entire category back at the end of 2015. Did BBC get DDOS'ed? I don't know. Look at the geos. This is what the geos look like. Pretty flat down at the bottom, but at the beginning of 2015, a giant spike. And guess what that was? That was Cymru and Wales. If you take them out, it looks

like a pretty smooth trend, but if you weren't looking at this to figure out what happened in these sections, you would think, "Wow, the category is quite erratic."

So, let me just summarize here. I know this is a lot of data, but I personally believe this open data initiative is really important to ICANN. I think it's really important to help us run our businesses. I think the tools that we've talked about here are really ways to help get at this data, and I think the data can help show adoption and use, not just the traditional, "How many names have you got in your zone?"

But, we have to use the data with care. There's a lot of large anomalies in here. Registry operators are not all reporting accurately, and nobody at ICANN is looking at this data to say that, "Hey BBC, when I got your last report it looked like you had a huge spike. What happened there?" And they would probably go back and say, "Well, sorry, that was incorrect. We've now looked at the data. One of our counting servers was off, and there was an anomaly."

But if we don't look at those kinds of anomalies and fix the underlying trend so we understand it, we're not going to get a real reading, and it will be a garbage in/garbage out kind of a data thing. So, I think the registry operator reports need a lot greater scrutiny. Yes, we need data tools, but I think we really

have to make sure that the data itself, whether that be registry operator reports, SLA reports, all the other stuff that ICANN is going to be much more transparent about going forward, we really have to be very clear that the data has to be accurate and complete and timely before we start making business decisions or policy decisions based on it. So, I think that's it. Thank you.

ALAIN DURAND: Thank you very much, Roland. Keep the microphone over there. So, I think we have a question in the online tool. And then we will move to questions in the room.

CATHY PETERSEN: Actually, we have two comments in the remote chat. And we have one raised hand in the room. Can I share the comments first? From John McCormac from HosterStats.com, his first comment was, "It seems that there is a combination of end users and Big Data people here. The tools are nice for the end users, but Big Data people may be more interested in raw data sets."

And his second comment is, "The ICANN registry reports from the ICANN site are generally quite trivial to integrate. The problem is that they don't have a standard format before 2009." Now we have the first raised hand in the room.

ALAIN DURAND: Just before we go to the raised hand, if you would allow me. I would like the panel to react to the first comment about end user versus Big Data, if you would like to share a few words. Anyone who would like to take this.

JAY DALEY: This is Jay Daley. Conceptually, you need to look at this on three separate levels. The bottom level is just the data that is made available, and data scientists will generally take that data and use that data quite well, quite happily. The second one is a set of more user friendly tools to visualize, to create simple charts, to do simple things with the data. The product we use, say does one and a half of those layers, but there are other tools out there that nicely add that layer on. Then, the third layer is the one of narrative. That's where you take the charts, you take the data and you tell a story about it and you do a deep dive into a particular thing. Again, there are separate tools that help do that.

So, I don't necessarily agree with Jojn that this is so focused in one way or another. As far as I can tell, the open data initiative understands all three layers of that and is looking at that, but it's impossible to find one tool that does all of those three things,

and so it will take time to build a layer of tools in place to support that whole level of interaction.

JONATHAN ZUCK: This is Jonathan. I would argue in favor of prioritizing the raw data because there are so many tools that are available in the marketplace for data analytics and that splitting the effort up now between raw data and end user experience tools, I feel like will undermine the rate at which this open data initiative can grow and that I would worry about providing tools much later when we understand how people are using the data.

CHRISTA TAYLOR: I think we should always have the choice. I know I use R for certain things. I use Python for other things. I'm going to switch back and forth depending on what I'm doing. The other aspect of it is I want to be able to choose what data is in my data set, and I want to make sure that I clean up properly, that's in the best interest of what I'm trying to do.

If I'm going to take everything within three standard deviations, I want to know that I have the full three standard deviations and not a set that's already been cleaned for me and that I don't see that data. That's not going to help me, and sometimes as I said earlier, it's the outliers where you get a lot of insight into things,

and to our businesses and those implications, so to take them out, I don't think is correct.

ROLAND LAPLANTE: Well, I'm a marketing guy, so I start with a narrative and then I look for data to support it. So, I really would object to getting the narrative from ICANN on the data, getting the report from ICANN on the data. I think we should just get the data in its rawest form possible and let us come to our own conclusions.

CHRISTA TAYLOR: One last comment on that. As a data scientist, you're going to come up with your own narrative after you run all your programs, after you look at all the statistics, and those statistics, if you go through those formulas, they're all different and it all depends on where your end point is. So, to say my narrative would be perhaps very different from Roland's and we're all going to have different perspectives on what we want from it and the tools that we utilize to get that end result.

ALAIN DURAND: Thank you all very much. Now, let's move on to questions.

MICHAEL KARANICOLAS: Hi, so my name is Michael Karanicolas. I'm the president of the Right To Know Coalition and the candidate researcher for the Open Government Partnership. So what that means is my day job is to design and assess and implement different transparency systems, usually at the governmental level, but I'm also the rapporteur for the CCWG Group on transparency.

So, that includes recommendations on a bunch of different initiatives that ICANN is taking forward including the DIDP, so whoever mentioned redacting the information or understanding the sensitive information, that presumably is on the DIDP characteristics which are going to be changing, so stay tuned for that. I'm also on the EC of the NCUC, so we're not all crazy extremists, but I appreciated the shout out.

So, congratulations on this very valuable initiative. I'm a strong believer in open data, and I think that it is of critical importance to enhance your transparency at ICANN. I also want to say that it's really great that you're starting with this census and overview of what information is available. Data management is the first step and understanding the scope of what exists out there is an important first step, especially because, as I'm sure you know, it's much easier to bake transparency into the system rather than trying to impose it at the end of the day.

So you talk about a streamlined approach where information is tagged as it's created, potentially is sensitive but with relevant keywords to enhance searchability. That kind of metatagging has to be done at the outset, because you can't do that at the end of the day. I'm sure that's not news to you, but it's very good that you are starting with this overview of what the information is because that is the best avenue forward.

I also want to say, just picking up on what was being mentioned here, that it is important to take a user-centric approach that reflects the broad cross section of people that are going to be accessing the information. So you hear, "Well, I'm the marketing guy, I want this. I'm a data scientist. I want this." There's going to be academics. There's going to be journalists. There's going to be people that are using the information for all different reasons and it's important to design the delivery of that information in a way that is user centric in its approach.

And one of the major problems that we see with open data systems as they move forward is accessibility and contextualization because as an organization of ICANN's size begins to put more and more information out there, it's very easy to get lost. People already complain about the website and the difficulty in coming there with a query and finding the appropriate data set. That kind of design, that kind of

contextualization and accessibility is fundamental to the design of the website.

So I wanted to with that very long prelude, I wanted to ask two things. One is whether you've given any consideration to adopting the open data charter which is a set of principles that has been endorsed by 17 countries and 35 subnational governments on the release and organization of information. I think ICANN would be the first intergovernmental or multistakeholder international organization to sign on to that, but they also provide resources for developing and implementing these systems, and I'd be very happy to connect you to those people, so I just wanted to ask if that's on your radar screen, and if not, hopefully you'll look into it.

And I also wanted to ask specifically about approaches to open contracting. That is an area that we've looked at a little bit as part of the CCWG. It's not an area that we very specifically pin down in terms of the avenue forward because it's kind of tangential to our mandate, but in terms of eprocurement and in terms of building greater transparency into the contracting system at the different levels, including assessments of the success and failure of contracts, this is a very important mechanism for accountability. So I wonder if that's something

that you've considered yet, or if it's on your radar screen. Thanks.

ALAIN DURAND: Suzanne or Marc, do you have a piece of information to share with us on this?

SUZANNE WOOLF: Actually, I want to make sure that I talk to you before you disappear because I am just starting the work of looking into what the procedure should be for determining the status of specific data, so the open data charter particularly, I really want to hear more about that. Thank you.

ALAIN DURAND: Thank you. Next question.

MARK DATYSGELD: Good morning. I'm Mark Datysgeld with the Business Constituency. I have been following the mailing list. I have been following the pilot. And I find it very interesting that there is a focus and I think it is quite important, but at the same time, I developed my Master's dissertation on the subject of ICANN PDP, and particularly looking at the GAC. And what I found out

there is that open data is a very relative thing, but the data being open doesn't exactly mean anything. It's not structured to reach firm conclusions on the data I was able to leverage.

It took a lot of processing from my part, very individual, and there is a lot of human data on ICANN that doesn't exactly relate directly to numbers, and long term, not thinking about immediate applications of the ODI, but has any consideration been given to starting to look, even if it's mid-long term, towards ICANN PDP, towards the efficiency of working groups, towards a more human-focused approach which at the end of the day it's the people who drive ICANN forward. So that is my question to any of you who would like to give me some sort of insight.

ALAIN DURAND: So, Suzanne, I'm going to put you on the spot also.

SUZANNE WOOLF: Sure. Yeah, this is Suzanne again. Actually again, I really like that question as we move into the next phase. Speaking only from where we've been so far with the census, I think one of the interesting issues and one of the reasons why we're starting with a working draft and we'll be revising it has to do with exactly the question you're asking about what's in scope? What's a data

set? What collection of bits are we trading as data sets for the purpose of the census and the open data initiative?

I think from the work Marc is doing on the pipeline, what we're going to do is probably have to prioritize the more numerical and formally structured data first. It also comes into question of how ICANN wants to handle responsibility internally, which I don't think any of us is really in a position to address just yet. But, that strikes me as a long term but important consideration. Thanks for the question.

JAY DALEY: Alain, can I answer that?

ALAIN DURAND: Please.

JAY DALEY: At the last ICANN meeting before in Copenhagen, Jonathan and I together ran a high interest topic session about the importance of data, and from a community perspective, one of the things that we were looking at is how can we have open data be there to support evidenced-based policy making because there is simply too much policy making that goes on that is based, as Jonathan was explaining on the CCT side of things, based on

people's supposition rather than actually looking at the evidence.

So, I think there is a push in the community that is not yet well defined and doesn't yet have ICANN support, as in resources behind it, but it is reasonably strong that we should be aiming for evidence based policy at some point.

ALAIN DURAND: Thank you. So, we have essentially two minutes left, So, I would like to take the two questions and if we can keep it really short so that we can have short answers.

JOTHAN FRAKES: I'll yield just after thanking ICANN. This is a very very important initiative. It helps many different things. My name is Jothan Frakes, I'm with the Domain Name Association. This information really helps fortify fact-based dialog rather than subjective, anecdotal things and it's been very very helpful as we convey a positive message and accurate message about marketing, and we have some of our members here in the audience. So, I wanted to thank ICANN for continuing this and encourage ongoing efforts towards this. Thank you.

ALAIN DURAND: Thank you very much.

DANIEL DARDAILLER: Hi, Daniel Dardailler, W3C. So I applaud this initiative of course, and at W3C we've published quite a bunch of standards for open data. One of the recent ones from January of this year is called "Data on the Web Best Practice." It has about 30 guidelines about exactly what you've been talking about: license term, use of vocabulary, all sort of things related to the persistence of URL etcetera, so I encourage you to go look at those guidelines. They are very important for what you are doing.

ALAIN DURAND: Thank you very much. One last comment from the panel and I will close.

JAY DALEY: So, I just want to add one more thing in response to the longest question ever asked at an ICANN meeting from earlier. I know that ICANN has looked at the open data charter because I recommended that they do that. So, it's not a not been thought about. Certainly there is consideration underway about that.

ALAIN DURAND: So, let's just go around for one last comment from the panel. Christa.

CHRISTA TAYLOR: I don't really have any last comment on my side. I think it's a great initiative and I really think it definitely should go forward along with the transparency and make sure that it really does gain some momentum. Thank you.

ROLAND LAPLANTE: Yeah, the only point I would make is to reiterate. I think we all say data is good. I just want to make sure that we're making decisions and policy and so forth based upon correct data and not anomalous data like we've seen in some of this.

JONATHAN ZUCK: I don't know that I have anything in particular to add except, you know, write your Congressman. No, I guess that's the wrong -- but obviously, data is incredibly important, and as the amount of what is going on at ICANN is exponentially increasing, the importance of being data driven in our policy development I think is equally imperative. When we had 20 gTLDs, then kind of shooting from the hip and guessing what would be a good idea

based on anecdotal evidence made a lot more sense than it does today.

ALAIN DURAND: Thank you all very much. I would like to give a round of applause to our panel. On behalf of Ed Lewis, thank you all very much for being here today. See you at the next meeting in Puerto Rico.

[END OF TRANSCRIPTION]