SAN JUAN – Cross-Community Session: Open Data & Information Transparency at ICANN: Developing Avenues Forward
Wednesday, March 14, 2018 – 10:30 to 12:00 AST
ICANN61 | San Juan, Puerto Rico

MICHAEL KARANICOLAS: Hi, everybody. So we're going to get started with the session on open data and information transparency at ICANN. My name is Michael Karanicolas. I'm with the NCUC, and I'm also the Work Stream 2 rapporteur for transparency.

And I work on transparency as part of my day job as well which is why it's very cool to be doing this -- having this conversation here at ICANN.

The format for the session today is going to begin with an introduction to the issue, open data, its impacts, its benefits, what we're talking about, followed by updates from ICANN's Open Data Initiative and Information Transparency Initiative followed by a couple of commercial -- couple of perspectives from the NPOC and the business constituency for their own perspectives on ICANN's progress and what the community should be looking for.

So thanks to all of you who joined us today. It's unfortunate that remote participation is going to be limited for this session, although on the bright side, I guess Adobe Connect being down,

presumably that means nobody has a reason to be on their laptop. Anyway...

So thanks to all of you for joining us, and we can -- oh, I can advance this myself.

So we're here today to talk about transparency, and transparency is the lifeblood of accountability. It's how we carry out our oversight. Everything that we as the community doing -- everything we do as a community should ideally be from an informed perspective, and access to timely, accurate, comprehensive and accessible information from ICANN is core to out ability to engage on the issues that the organization faces.

Now, when we talk about transparency, fundamentally what that means is a culture shift from a traditional model where information is classified by default and you're shown a limited amount, which is selected by the organization, to a model where things are open by default, where all information should be by default accessible.

Now, we're not talking about total transparency. There will always be information that needs to be kept confidential for reasons -- for legitimate reasons such as personal privacy or information whose disclosure would undermine measures to protect the security and stability of the Internet.

ICANN
COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

But, fundamentally, we're looking for a system where access is maximized. Access and openness should be understood expansively and exceptions to disclosure should be understood in as limited a manner as possible.

So we're going to be talking about two different avenues for disclosure today which are via the Open Data Initiative and the Information Transparency Initiative. We're going to learn more about what specifically these two programs are doing. But for now I wanted to sort of introduce the topic by giving a few examples of how open data can be effective and the impact that it can have on institutions that embrace transparency.

And I'm going to start by talking about transparency's economic benefits because there's been a lot of talk recently for the need to cut costs. It's a natural question, what kind of impact open data may have on an organization's bottom line. Experience suggests that in addition to external business benefits flowing from the information that ICANN can provide, open data at least in the form of procurement transparency can actually have a significant potential to save the organization money.

What you're seeing behind you now is a screen shot from Paraguay's open contracting website which was first introduced in 2010. Unfortunately it's just a screen shot so I can't show you all the cool functionalities and bells and whistles. But if you're

ICANN 61
COMMUNITY FORUM
SAN JUAN
10–15 March 2018

interested in seeing it further, if you look up Paraguay DNCP, you can go to the website and play around with it. It's very interactive. It lets you track who's spending how much, what's being spent on different costs. It even lets you drill down into specific individual contracts that the government is signing with different procurement entities.

I'll show you another website which is similarly very nice. This one is from the Ukraine. It's called ProZorro. Again, all the values that you see behind you are going to be clickable. You can dig down. You can find the individual contracts. You can see what's being spent in different time period by different entities.

Now, the costs for developing this website as a pilot in 2015 and then operating it through the end of 2017, were about -- 4.69 euros. But the cost savings in procurement as a result of open data -- of the open data system over that same two-year period can be estimated at $1.19 billion by calculating the costs that were actually being paid against the cost estimates that were derived before the procurement process started. So Ukraine basically, $4.69 million to build and operate the website. $1.19 billion in cost savings over the same time period. So those are just for cost estimates.

What about for actual spending? Recently I carried out a study for the Open Government Partnership looking at procurement costs in Paraguay after the adoption of their own open contracting system. Spending on things like defense and education are variable so we looked at prices paid for budget items which are relatively resistant to fluctuation. And if you look at the chart behind you, between 2010 and 2016, the government saved $150 million on office supplies alone after the introduction of the open contracting system.

Now, these are just some examples. It's uncertain if or how openness in procurement and contracting will be a part of these initiatives. It is a dataset which could potentially be part of the open data approach, and hopefully it is being considered.

Before I move on from the Paraguay example, I'll also mention that transparency is very useful to uncovering waste or mismanagement. After transparency was adopted in Paraguay, journalists and citizens began to dig through the books and found $35,000 in spending on refreshments for a single meeting by the ministry of education. So this triggered widespread protests.

The ones you see behind you, it's called the Cocido de Oro protests for gold and tea because they were spending $14 for single pots of matcha tea. And ultimately it led to the

ICANN 61
COMMUNITY FORUM
SAN JUAN
10–15 March 2018

resignation of the ministry of education as a result of these protests.

And, again, it's the impact of that oversight and public accountability which not only can lead to -- which not only can lead to proper accountability for mismanaged resources but can also promote -- but can also serve to promote a culture of responsible spending.

Give a couple more examples before we turn it over.

This is Canada's open parliament website. You can chart what's being said by who in what context. Accessibility of information is a major challenge for everybody here. As I engage particularly with new members, I constantly hear about frustrations about how difficult it is to be brought up to speed on the issue that they want to engage in. This can be a problem with veteran members as well, where you step away for a few weeks, come back, and suddenly the conversation is in a new place and it's very difficult to chart what's been going on, very difficult to chart to get sufficient background of the debates that are going on. So open information, open data can be key to delivering that information.

This website, for example, allows you to search for any term and find how it's been raised in debates in the Canadian parliament.

So, for example, if you want to look at who's been talking about steel and aluminum in the aftermath of the announced tariffs that are going forward on Canadian steel and aluminum, you can find it right away. So again, information accessibility, I think, is going to be a big part of this discussion that we're going to be hearing about shortly.

I'll wrap up with a few more examples from the U.K. One of the cool things about open data is it's impossible to predict where the community is going to take things once the information gets released. So the U.K. has put up a website called data.gov.uk where they are basically just releasing these datasets people are free to use.

And you see it being used in all sorts of innovative ways. This is a map of London with traffic accidents charted out so that you can see which intersections are most dangerous. You can specifically look and see where is it most dangerous to approach as a cyclist or as a pedestrian, maybe plan your route to work, maybe plan where you're going to slow down. Lots of very useful stuff.

These are statistics showing arrests and banning orders organized by football club. So you can track where hooliganism is worse and which fans are doing stuff more than others. There isn't a fancy chart to go with this but I will just tell you, New

Castle is way out in front, but Chelsea made a very strong showing as well.

And bringing it back to Puerto Rico, this is a map of accessible establishments that's being used by again integrating information from public databases on which establishments are accessible to people with different disabilities. And you can plot that on to a map of where we are right now and get all sorts of information that's useful for people trying to get around.

So with that kind of brief introduction to open data, I would like to turn it over to Jay Daley who is a community member contracted to work on the Open Data Initiative who has been active on open data issues for a long time. And he can introduce the Open Data Initiative and talk about his progress.

JAY DALEY: Hello. My name is Jay Daley. And we're just going to try to get the other set of slides up for me.

So I'll fill in for a moment then. Michael has given a great introduction to the Open Data Initiative -- sorry, to the benefit of open data generally. So I'm not going to cover too much of that, but I do want to explain what the Open Data Initiative is and the specifics about it and particularly because there are two initiatives going on at the moment, the other one being the

ICANN
COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

Information Transparency Initiative, which you'll have a presentation about in a moment, just to be clear about the difference between the two of them.

So the Open Data Initiative is about enhanced access to raw data. Now, by "raw data," we generally mean tabular data, the type of data you might find in a spreadsheet or a database that is normally machine readable and that a data analyst or a marketing person or somebody else may wish to pull into a tool and then use within that tool.

The intent here is to get that data moving, get that data flowing, get it being reused because the more it's reused, the more value it's got from that data for the whole community and for the organization.

Doing that will engender trust within the ICANN community. As Michael says, this is an overall part of a number of projects that are aiming towards transparency.

And one of the key points that has been identified by ICANN to come from this is enabling evidence-based decision-making. As a number of you have seen on previous discussions about this, there are multiple misconceptions out there such as that registrants prefer very short domain names and yet all of the

data shows that the medium size of a domain name length is about ten characters long that people prefer.

So just to be clear, the Open Data Initiative is not the same as the Information Transparency Initiative. That will come later, and it's not duplicating effort.

So the process then, Goran Marby has taken a very strong personal interest in this and made it one of his goals for some delivery by July the 1st. And that's a strong public commitment to open data.

We have also now launched a tender, a request for proposals, to acquire a software as a service open data platform. That was launched on Monday. And so -- and this is a specialized tool that is intended to enable the reuse and the dissemination of open data, particularly through open standards so that anybody can access it from any open tool.

It will take some months to go through the process, and there are several well-established vendors in this market plus many other people trying to enter it. So this should be a competitive process.

Now, before reaching this point, a number of open data pilots were put together using tools, which I'll explain a bit more on my final slide.

And that gave a chance for those community members who were familiar with this to work with those tools to download datasets and to see how that worked. And that's now complete, and feedback from that has been incorporated into the RFP.

So internally the work has been on something called a Data Asset Inventory. So this is a list of all the datasets within ICANN and various attributes about them, such as whether they contain personal identifiable information, how that needs to be redacted, details about how they're stored and things. And this has now been -- we have the framework in place, and this has now been filled out through a process of interviews and other research.

The intent from this is to provide a public data catalog which can then be released to the community so that we can then get your feedback on the prioritization and the importance of the datasets within there.

The process for release is more complicated than just the priority order because it also has to consider how difficult it is to be able to get certain datasets out. And so there needs to be a trade-off between the two, but the community views will be exceptionally important on that.

And so that will be released relatively -- by the time, we hope, July the 1st. And from that, those are the datasets that will then go on to the open data portal.

And then, finally, internally we're establishing an internal data governance model around these datasets, around this Data Asset Inventory.

So, finally, just to recap on the Open Data Initiative pilots, there were four of these: CKAN which is an open source tool used by many governments. It's a very strong cataloging tool and many governments have used it because of that. And then there are three commercial providers: OpenDataSoft, Enigma, and Socrata. A number of datasets are available on each. Two of those is still available for you to look at.

So that's it for me. I shall hand over back to Michael. Thank you.

MICHAEL KARANICOLAS: Thanks so much for that update.

And we can now go across the table to Jana Juginovic who is here from ICANN staff to update us on the ITI.

ICANN
COMMUNITY FORUM
61
SAN JUAN
10–15 March 2018

JANA JUGINOVIC: Thanks, Michael. We're here to talk about the Information Transparency Initiative.

I know some of you may have read about did on a few blogs that we posted on icann.org, and we're going to tell you a little bit about the progress that we have made so far since we launched the project in January of this year and since it was approved by the Board in September of 2017.

The heart of the Information Transparency Initiative is to increase the accountability and transparency of our information, and icann.org has made those commitments. And this is in, as you had mentioned about ODI being one of Goran's goals as is ITI.

Current and easy to find information in all six UN languages is an important part of meeting those commitments, and it's one of the ways we demonstrate our transparency and our accountability.

How are we going to do this? So the first thing that we need to do is to develop a coherent and consistent content strategy. And to do that, we're going to be building a comprehensive taxonomy and information architecture and a content governance that we're going to apply to our external content.

How we're going to do that is to deploy a unified Document Management System. This is at the heart of this project. For the first time, ICANN is going to be implementing a Document Management System and building that foundation, and with that we're going to be able to implement that taxonomy and that information architecture.

We're going to surface that content through a new singular, scalable, secure Content Management System. So that is how our public information will then be displayed to the community.

We have nine goals, which you can read about in the Board resolution that was passed in September 2017 when ITI was approved. I talked about some of those goals just previously, about building that taxonomy and information architecture, and the Document Management System and Content Management System. But some of the goals also include establishing a future-proof and content agnostic technology landscape. That means that if we have a new Content Management System, new technologies evolve, that that content -- we don't need to change it because we've already applied taxonomy to it, so we'll be able to easily move that over to a new Content Management System.

Through this new infrastructure that we're building, the DMS and the CMS, the Document Management System and the

ICANN 61
COMMUNITY FORUM
SAN JUAN
10–15 March 2018

EN

Content Management System, we're upgrading our technical infrastructure and thereby able to serve our global community better through increased findability and accessibility of our content.

I want to talk a little bit about the benefits that this is going to do for the ICANN community and the organization. You know, we've heard a lot of feedback from the community that they have difficulty finding content on icann.org. So the first thing that the Document Management System is going to tackle and ITI is going to tackle is our public information.

So the first benefit that we want to have and display for the ICANN community is to improve the findability, transparency, the security, and the quality of our public content.

One of the other benefits, and we've had a lot of discussions about, obviously, the budget here at ICANN61, is doing this is going to decrease yearly maintenance cost for our -- for our content because it consolidates 15 separate sites that we currently manage. And a lot of the way that we get content to the site is going to be by staff, so self-publishing, which will decrease cost without relying on a web management team. Again, I mention it's going to ensure that our content is future-proof because we're going to apply a taxonomy to it, so if we do

ever need the content to a new system that that taxonomy is there and present forever.

It lays the foundation down for an ICANN shared ecosystem of the DMS and CMS as Document Management System and Content Management System that we'll be able to apply to all of our SO/AC sites once this foundation is laid out. It's going to improve the accessibility in all six U.N. languages, and it helps us meet those commitments and reduce financial risk.

In terms of what we've done so far, since January, we've had a lot of activity. We -- Again, I mentioned, we started in January of this year, and we've started by tagging all of our content. We have a lot of content on icann.org because as I mentioned, that's what we started with. We have about a 104,000 pieces of content. That's a lot. And we've started tagging 9,000 pieces so far which are done, and we're doing it with a combination of eyes-on audit, so ICANN staff applying taxonomy directly to that content, and through machine learning, which is able to recognize certain patterns in language and apply consistent taxonomy to that.

We've already built the technological infrastructure, the DMS and the CMS, and done the initial integration between the two platforms, so that is complete.

We've conducted interviews with the ICANN community, with the chairs of the supporting organization and advisory committees, regional leaders, registry and registrar operators, NextGen and fellows, registrants, and other stakeholders.

We'll be launching a feedback.icann.org site, and this is really important because this is where we need the community to tell us what they want ITI to be, how they want that content to be surfaced, how some of the features will be rolled out, like public comment. I know some of you have talked about that. So we'll be launching these features on this feedback site and content and give you direct visibility to what we're doing and to give you the opportunity to provide that feedback directly because we need it to work for you.

And we'll be -- the first thing that we're actually going to be testing on this integration between the DMS and the CMS is the new ICANN acronyms and terms to replace the glossary. And I know some people have trouble recognizing ICANN acronyms and terms so we're hoping this feature will improve that.

And that is all we have. This is what the feedback site is going to look like when it's launched, and we'll be sharing that information on a blog on icann.org and through social media and let you know when you can start accessing that feedback site.

ICANN
COMMUNITY FORUM
61
SAN JUAN
10–15 March 2018

Thank you.

MICHAEL KARANICOLAS: Fantastic. Thanks so much for that. And feedback is incredibly important. That's one of the biggest reasons why we're here.

So why don't we hear from a couple of people from the community, starting with Raoul Plommer who is a member of NPOC who works at Open Knowledge Finland. And then we'll -- after a couple of more presentations, we'll go out for direct feedback from the community here.

So, Raoul, do you want to take it away?

RAOUL PLOMMER: Thanks, Michael. Good morning, everyone. Name is Raoul Plommer, vice chair of NPOC, which is a constituency of the NCSG, representing civil society under the GNSO.

Because my member organization, the Open Knowledge Finland, is specifically focusing on increasing transparency and openness in the world, specifically in Finland, I was asked to do this presentation, which after a fair bit of scrambling formed into mere commentary of the ICANN ODI. If nothing else, done some work in collecting information to this presentation, and that can

ICANN
COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

be found at okf.fi/ICANN-ODI. That's -- the latter part is a lot of capital letters. So ICANN hyphen ODI.

So can go to the next slide, please, or -- Thank you.

So the ODI international is a growing international efforts with groups in more than 40 countries. And everyone is welcome there. Anyone can join the network, individuals and organizations, and you don't really need special skills or background. I guess that's how I got in there.

These are the -- lists of the -- of the tools provided to us. And I just tested them this morning, and the CKAN doesn't actually resolve. Took too long. I don't know, that could have been my connection, but I doubt it. Did I try it with different browsers. And on the website of ICANN, it said that it's only Enigma that needs registering, but also OpenDataSoft needs that.

The last one, Socrata, you could actually access immediately and see what's there. Yay!

Now what is in Socrata? There's only two data sets, updated on the 28th of June 2017. I did read elsewhere that these would be updated, like, three months in arrears, but, like, that's quite a bit longer than that.

And like this presentation, if you actually find the page, all the -- they all provided links to these. Like, basically everything that's underlined are actually links to those pages.

The only one I was really being able to understand is the external data set of 1263 gTLDs, and the three columns that actually made sense to me were the date of contract signature, the name of the gTLD, and the name of registry operator. Now, I couldn't -- I couldn't really make too much sense of that.

Now, it's -- it's good to measure the success of opening data, and I think that's a -- that's a really important part of taking this project on. The number of data sets is a bad success measure. It incents the wrong behavior. If we publish every year as a different data set, our number goes up but data usability goes down. If we publish low-value data, our end goes up, but no one wants to see it.

So there are key performance indicators that can be used to evaluate the usefulness, and those help us to stay on track with the implementation of the Open Data Initiative. So we should really be using that to be able to evaluate it.

They help us in identifying and communicating our successes and flag areas where we might need to invest more.

To measure the success of the initiative, there are basically three main -- main things. It's basically the readiness, what's the capacity to take on the project and to deliver it; implementation, what are the key data sets being published and made available; and impact, like real-world results of the project. Those could be, like, new products and services as well as insights.

Also, it's good to have a demand-focused approach, which would actually prioritize the needs of the data user over the data publisher.

Now, you may ask, like, what's the value of this project? It's not going to be exactly cheap, but I did find a good list of value, like, where the value creation actually happens. And I think these are all pretty good things, valuable things to strive for.

Now, open data creates opportunities to increase the level of data literacy in the workforce by raising public awareness of data skills. Opening data can encourage people within an organization to communicate with each other and with external groups.

For businesses, opening data can improve relationships with consumers and open up new customers channels.

Open data has the ability to bring about change within societies. Examples of open government budgets have been very useful.

I think it would be useful to have open data of ICANN's budget as well.

Open data helps communities to communicate with each other. There's this project called "My Society" who created a "Fix My Street" which allows residents to share problems with their local areas, open data, and create shared solutions.

Now, these are suggestions I came up with. We -- we really need to know, like, what data ICANN has. It's been really surprisingly difficult to actually find that out. Now, I did hear Jana said it would be open -- or, sorry, Jay. Jay said we would have a report of the inventory, at least a preliminary one, on 1st of July. So I'm quite keen on seeing that.

We -- One of the functionalities that was also mentioned in the ICANN web page was to give users the ability to receive new public content directly to their inbox if that's their preference. And we could also, instead of just using the data from the registries, we should also be using it to track the effectiveness of policy. And I've listed a few -- a few metrics that are used by governments, which are namely the International Open Data

Charter, Open Data Index, and Open Data Barometer. So we don't really have to reinvent the wheel there.

And conclusions are that, yep, to promote this initiative, initially I was told that -- that this project really needs a push from the community. Now, I think for us to do that, we really need to know what's there. So that inventory will be important in doing that.

And I think we should also ask the community what should we prioritize on.

That's it. Thank you.

MICHAEL KARANICOLAS: Thanks so much for that, Raoul.

So now I will pass it over to Mark Datysgeld who holds a master's focused on Internet governance and helps represent Latin American business within ICANN's business constituency. Mark.

MARK DATYSGELD: Thank you so much, Michael. Good morning, everyone.

So, actually, coming from the perspective of the business constituency, the Open Data Initiative has been something that has been pushed forward, both in terms of budget and in terms

EN

of a need for representation by the business constituency in previous comments. So the call, the main call, has been for transparency as far as domain abuse is concerned and as far as how the investments of the domain system is being made. But at the same time, something that I find interesting is that we, as a community, have not yet come together to understand what we want out of the Open Data Initiative.

To me, it seems very prioritary that we look at domain abuse and that we look at domain abuse and that we look at the way the domain industry is being shaped because this is the core business of many people who are in community and it also serves to form policies, but at the same time, I do think that there are many opportunities that can go beyond that. And this is where I come as a researcher.

Back when I was doing research for my master's on the subject of ICANN and policy-making, I found that it is actually incredibly difficult to track how policy is formed in ICANN. Sure, we do have open data in the sense that the data is open, but what I found is that open data is just that. It's data that is open. If you don't have a system to properly catalogue, if you don't have something that informs the way that it's being used, if you don't understand what you want out of that data, it will keep on being a pile of data.

So we need to come together and start thinking about what we want. And this is what I thought lacked in the first implementation of the ODI initiative.

Sure, we were aware it was going on, and we did interact with those who were conducting the process, but at the same time, there was not a sense of what exactly does the community want out of this. Where are we trying to get? What we thought at the time is we don't know what data sets are available; therefore, we cannot know what we want. But at the same time, that seems a bit self-limiting; right?

So I do think that, moving forward, there are different aspects that we can look into as far as what we want open data to be and what we think open data can be.

So, for instance, from my perspective, from my research, I would very much like to understand how regional participation goes on within the PDP. I would like to be able to examine that in a broader manner, for instance. And as well as stakeholder engagement. When we are doing outreach in the LAC region, how is the engagement of people from different regions being carried out in specific manners? Those are things that go beyond the domain name industry or indicators that ICANN org can provide. But at the same time, they are data that is made

available on ICANN and that we could make good use to produce policy.

So I guess my key take-away from having engaged with researching ICANN and trying to use its data to produce meaningful research is that we need to establish a better focus of what we want. And carrying forward with this initiative, now that we have the budget commitment, we have the staff commitment, we have a community forming around that, it might be of our best interest to start discussing before our constituencies and before we are working with, like, what are you expecting out of this? What would you like to know? Because the data is there. It's -- The problem is that it's -- it's -- it's too divided. It is between PDFs and Excel spreadsheets, pure HML and/or text files or something that's in an obscure FTP somewhere.

So even those resources that seem a bit out of place can be put better to help us create meaningful policy. And I guess as my message, we as a community can make this into something bigger or we can settle for it being a data analysis, too, for domain names which in itself is incredibly meaningful. Let's not reduce the value of that. We need to understand how this is being shaped. But at the same time, we can take this a bit further.

ICANN
COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

I do have some -- some more points. But at the same time, we're running incredibly late and I think it would be very good for us to provide opportunities for the audience to engage. So I will pass the mic back to Michael.

MICHAEL KARANICOLAS: So thanks very much for that. I congratulate all of our speakers on adhering to our tight time frame which will allow us to have a good amount of time for a community discussion.

So with that being said, we would like to invite any commentary, any ideas, anything that anybody wants to share, questions. Please feel free.

MASON COLE: Hi. Good morning. Mason Cole speaking. I'm curious -- I applaud ICANN collecting as much data as possible. But in order to avoid what you talked about before, just having a big pile of data, how will you also organize the data internally so it can be applied to ICANN functions? For example, will some of the data that you've collected be applied to the compliance function or to other functions internally inside the ICANN org?

MICHAEL KARANICOLAS: So that's an interesting one. I guess, is that more on the ITI side? I guess the question could be -- if I understand correctly, is basically will there be applications within ICANN as well?

JAY DALEY: One of the interesting side effects I have seen of doing an open data implementation is that the data becomes more accessible to people within the organization as well as people outside of the organization. Because there is now a single catalog, there is a single point to get the data from and because that data is guaranteed to be updated regularly.

And so, yes, it doesn't actually lead to that certainly. As to whether there are specific plans in place, not yet. But that is well understood.

MICHAEL KARANICOLAS: Great. So why don't we go over to Number 3 next, and then I think we have a remote question after that.

MONA ELSWAH: Hi. I'm Mona Elswah from Egypt. Thank you for the initiative and for -- and I think on behalf of every data scientist, we need to thank you for this. I think we will see after the initiative is released many research on -- using this data.

ICANN COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

Just wondering if there are -- are there any constraints on the datasets? For example, are some data will be anonymized? Are some data will not be released, or everything would be available and free for people? Thank you.

JAY DALEY: So, yes, there will be some constraints. Certainly, there are legal obligations about the redaction of personal identifiable information. And there are other sets of data that are potentially so confidential that they cannot be released.

Now, wherever possible, it is our intent to still document those datasets and let you know about those. So that can be an open and transparent decision-making process. But I think it is inevitable some data will not make it through because that would be counterproductive or potentially legally harmful if it were allowed.

But there is -- to be clear, there is a very strong top-down culture of openness and transparency within ICANN. And I can say that as a longstanding community member now being able to look inside it. And so this is not going to be used as an excuse to hide data.

The default will be to publish, but there do have to be some legal exceptions that control that.

MICHAEL KARANICOLAS: Did you want to -- did you want to add anything on the ITI side in terms of tagging out?

JANA JUGINOVIC: So for all the information that we're tagging right now, it's all public information. But we are adding additional taxonomy around -- to comply with GDPR. So anything that, as you had mentioned, has any private information like an email address being tagged that's on the website, so we are compliant with it. So that taxonomy and that tagging process is taking that into account.

MICHAEL KARANICOLAS: Is that a remote question? Please.

REMOTE INTERVENTION: We have a remote question from Dev Anand Teelucksingh.

I've done my own ICANN stakeholder analysis tool in Google Sheets which attempts to analyze the ICANN stakeholder communities by country and by region or worldwide. The idea behind the tool was to, one, be able to find all the ICANN

ICANN COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

community members in a country or region; two, identify stakeholder gaps in countries and regions for targeted outreach.

So far the tool has the drill-down ability to search by country; one, At-Large representation; two, ccNSO membership; three, GAC membership; four, fellowships; five, NextGen; six, accredited registrars and also do a regional or worldwide analysis on At-Large, ccNSO, GAC, and accredited registrars.

It attempts to pull the live data from accredited registrars' Web page, ccNSO, and GAC website highlighting a possible quick win that public data from current ICANN websites can be scraped and reused if information from such community websites display data in a consistent manner.

A challenge to note was that different AC/SOs uses different names for the name country/territory requiring such country/territory names to be standardized across the AC/SO for comparison within the tool.

The link to the latest version of the stakeholder tool can be found by Google "ICANN Stakeholder Tool."

MICHAEL KARANICOLAS: Thanks very much.

I think I received an email about that initiative as well at some point, so I think it's been shared with at least some of the lists.

I do feel like that taps into this enormous potentiality that's there where I'm sure there is a huge community of researchers and data geeks out there who would love to have access to this and are looking forward to different opportunities to make use of it.

Mark, did you want to respond?

MARK DATYSGELD: Yeah. Actually, I am aware of this tool. It has been shared on the ODI list. And it goes back to the point I was making about how can we use open data in different manners. We are not aware of how our community is spread exactly. We know who goes to meetings. We can have a rough idea. But this tool is, like, a first step into understanding where exactly ICANN is happening not only in the meetings themselves because this is not a good sample of what the community looks like.

And I'm aware that at least somebody from the ICANN org side has been developing something along those lines. I could be wrong. But this is a project that could perhaps be somehow brought together of that.

ICANN
COMMUNITY FORUM 61
SAN JUAN
10–15 March 2018

MICHAEL KARANICOLAS: Great. But of -- Sure.

JAY DALEY: Just briefly to add to that, there is a well-known five-star deployment scheme for open data which has as the fifth star of it linked open data such that the inconsistencies that were highlighted there are eliminated because there is a central, very clear data dictionary about everything. And that is then linked in a machine-readable way.

We're not aiming to get that far yet. We're only aiming for three-star open data, but that will be looked at in future. It's just a significantly greater effort to get to that level.

JANA JUGINOVIC: One thing I wanted to add is some of that information of where the community groups are coming from for -- particularly for the meetings are available in the post-meeting report. So it breaks it down. People are able to identify what SO/AC, what group they belong to. So you are able to see it in the post-meeting reports about a month after each report.

ICANN 61
COMMUNITY FORUM
SAN JUAN
10–15 March 2018

MICHAEL KARANICOLAS: Why don't we move back over to Number 2.

MARK McFADDEN: Hi, Mark McFadden from the ISPs. This is a question for Jay.

I'm looking at a letter that the Commercial Stakeholder Group sent January 5th, 2017. It's about 15 months ago.

And in it, those three constituencies made very specific requests for access to open data and specific datasets. It was not -- it was not some vague request. It was actually very specific requests.

And in March -- a year ago, by the way. In March the CEO of ICANN responded that, indeed, those datasets would be made available as expeditiously as possible. A year -- more than a year has gone by, Jay. And ICANN as far as I can tell has made no further progress in making those datasets available. They were very specific requests for very specific purposes that those constituencies were interested in.

And, I guess, first of all, am I wrong? There has been no progress? There's been no public progress. For instance, there's no publicly available data for us to work with.

And the second thing is that if we can't make progress in a year, why should we have any confidence that the next year, 2018, will be any better?

EN

I'm sorry to make this guy talk. His throat is so sore.

DAVID CONRAD: David Conrad, ICANN CTO.

So we have been making progress. In particular, we've been trying to narrow down the platforms upon which we will be publishing the data. We have been working internally to identify the various datasets.

The three datasets that were provided by the business constituency I don't actually recall -- or ISPCP I don't actually recall off the top of my head. But we are aware of those requirements, and they are prioritized in the deployment once we actually get to the stage of putting the open data platforms into production.

One of the challenges that we have is that there is a very large amount of data within the organization, most of which is not in forms that are very amenable to deploying within a structured data system. And we're trying to identify those datasets and ensure that the data that is specified within those datasets is known, that we have sort of a dictionary -- a data dictionary that describes those datasets, and that we don't cross over into violations of privacy or contractual obligations for that data.

So we apologize for the delays in getting these datasets open and available. But it is something that we are obviously continuing to work on as shown by us bringing in Jay as a consultant on this particular project as well as hiring additional contractors internally to drive the project forward.

MICHAEL KARANICOLAS: Just wanted to follow up briefly?

MARK McFADDEN: I have to.

So thank you for that, David, and I appreciate that. And I think I can speak for many members of the community that we welcome having Jay on board to work on the project.

The request that I would make now -- I mean, we can't do anything about what happened in the last year except express the Commercial Stakeholders Group's regret that we didn't make much progress on visible data sources. But we understand the complexity that ICANN faces in terms of taking the data sources and relieving them of their contractual obligations, et cetera.

But here's my request to you and Jay at this point, is that as we move forward in the next months that the community get more

expeditious progress reports, that we actually get some visibility as to what's happening.

We saw the RFP come out earlier this week. That's great. But as Jay makes progress and as your team makes progress, one of the things that would be very helpful for our constituencies is to be able to report that progress back into our constituencies and show that we're making -- actually making progress on the ODI. So that's my one ask here.

DAVID CONRAD: We will commit to providing progress reports on a periodic basis.

One request that we would have is that we have to date received very little input from the community, including the Commercial Stakeholders Group about what exact datasets they would like, what features they would like to see in the open data platforms.

The two platforms that we have settled on, we have done primarily due to internal observations, internal requirements gathering. We have not received a whole lot of input from a community as to what they would like to see within the platforms. So that's another area that we would request additional help.

MICHAEL KARANICOLAS: I don't want this to turn into a two-person debate. So we do have questions all around the room.

Why don't we go over to Number 3.

CHYLOE HEALY: Hello. My name is Chyloe Healy. I'm from Canada, and I'm a newcomer. So I apologize if this is not an appropriate question, but I don't know -- because I don't know the datasets.

I was just curious, though. I know some datasets do need context or there's an increase of misinterpretation, which could be harmful.

I'm just curious if ICANN will play a role or have some sort of way to mitigate this.

MICHAEL KARANICOLAS: As a bit of context, Chyloe comes to us from the First Nations Internet Governance, FNIG, Center in Canada where contextualization of information has been an interesting debate going back and forth around First Nations' information that's held by the Government of Canada. Does anyone -- Please.

JAY DALEY: So in my experience of this, national statistics agencies are very concerned about contextualization of their data. And they tend to annotate the data to some degree, shall we say, to some significant depth in order to control the contextualization. And the normal activity of most data scientists is to strip that annotation immediately and discard it.

It is very difficult to put the contextualization at the data level. It is something that generally needs to come at a higher level in the process to be able to do that.

Now, because ICANN is an open organization with a firm commitment to openness, there's not going to be an attempt to contextualize that data in that level in that type of way as there might be with a national statistics agency, for example. And I'm unaware of any plans to do anything above that because I think this is a genuine desire to see how the community takes this data.

DIEGO CANABARRO: Hi. This is Diego. I work for the Brazilian Internet Government blah, blah, blah. Can you hear me?

MICHAEL KARANICOLAS: It's one, microphone 1.

DIEGO CANABARRO: Okay. Thank you. This is Diego. I work for the Brazilian Internet Steering Committee, but here I speak in my personal capacity. First of all, congratulations for the initiatives that were just presented. They are amazing for various specific reasons. But it struck me that none of you said anything about the W3C standards for this plain data on the Web and for structuring data, for creating infrastructure that will be used in display data online.

So I was just wondering: What role do the W3C standards play in your initiatives? Do they have any role? Yes or no. Thank you.

MICHAEL KARANICOLAS: Start with Jay, and we'll see if Jana wants to add anything.

JAY DALEY: I did in my presentation mention the requirement for open standards for access to the data. I just didn't go into the details on there.

The CSV on the Web standard is still very new. It is still, I think, a while off being properly supported by any vendor yet, okay? So we have in the RFP asked for details about open API compliance, adjacent APIs and about CSV on the Web compliance. But there

may need to be a pathway around this or trade-off. I don't want to say too much because the RFP is underway.

And as you probably know, we also -- Microsoft Odata is a significant industry standard even though it is a proprietary standard out there as well and built into a number of tools.

Just about all of the suppliers make the data available through multiple different mechanisms for people to achieve that. And, yes, I agree with you, the end goal of this should be entirely open standards in that way. But this is a maturing market, and so the time scales for that are yet to be seen.

MICHAEL KARANICOLAS: We're back over to number 2.

DENISE MICHEL: Hi, this is Denise Michel with Facebook. I think -- I want to thank you, first of all, for holding this session. It's a critically important issue, and we really appreciate that David Conrad and his team is making this a priority.

I would also note that the business constituency, which Facebook is a member of, recently submitted comments on the FY18 budget, strongly supported the ODI initiative and increased funding for it.

With that said, we have been providing detailed input on data sets for two years now. And so it seems a bit like Groundhog Day on some of this. And it would be great if the substantial input that many individuals and companies and constituencies have provided on data sets that will be useful over the last two years, literally two years, would be reviewed by the staff working on ODI and incorporated in your current work rather than stopping and asking for whole new sets of input.

We -- Many companies and organizations used -- had our scientists and experts provide very detailed inputs, sit with our staff for hours providing guidance and ideas for this.

I would urge you to use, please, that work over the last two years. And since the business constituency is supporting -- is asking for financial support again for this effort, we would underscore, I think, Mark McFadden's comment that we would, this time, like to see real deliverables and results and appropriate project plan with the timeline.

Thank you.

MICHAEL KARANICOLAS: Sure.

JAY DALEY:	So both Denise and Mark, certainly a commitment I can give that we will try to build that in so by the time we come to you, you know, it's not Groundhog Day for you.
	Back over to -- Nope. Another -- Again on number two.
FABRICIO VAYRA:	Hello, Fabricio Vayra from Perkins Coie. Just to start off, I do echo and share the concerns that both Mark and Denise have on the timing. I welcome everything you guys are doing and I really, really appreciate it. My question was going to be about the ITHI. And the reason that the timing issue is sensitive to me, I was at the original ITHI meeting September 7th, 2016. And while this isn't ITHI, I did see in the panel someone ask is that part of the ITHI.
	And so my question, aside from sharing the concern about the timing, is just what is the overlap between the two? Because I'm not entirely sure that the community is clear on that, and I saw a little bit of questioning on the panel itself. And how can we ensure that we're, one, not doubling efforts, and, two, if this is a Venn diagram, that we're capturing everything that's outside of the overlap in one effort so that data becomes open, because open sooner rather than later, and it's useful, and we don't have multiple tools that either conflict or bring up different things.

MICHAEL KARANICOLAS: So maybe it would be useful to talk about the integration between those two as well as maybe to reiterate the...

Okay. Sorry. Go.

JAY DALEY: So David may need to give a Fuller answer but -- No, Matt is going to give a fuller answer.

MATT LARSON: Hi, Matt Larson, VP of research for the office of the CTO.

So I want to just briefly talk about the difference between ITI [sic], which is the Identifier Technology Health Initiative, and the two things we're talking about today. So the ITI project is developing a set of metrics to assess the health of the unique identifier systems that ICANN helps coordinate. The RIRs are working on the numbers side of that, metrics for the numbers. In the office of the CTO, we've spent a lot of time developing metrics for the health of names and also, to a lesser extent, the protocol parameter registries. So these are all data that we're going to track over time, the idea being that you'll be able to see, over time, changes in these metrics will reflect the health of the identifier systems.

ICANN 61
COMMUNITY FORUM
SAN JUAN
10–15 March 2018

So this being data that we're going to track, it's the perfect sort of data set to go into ODI.

Other than that, other than the ITHI data being data that will be published on the ODI platform, there's no other relationship between the projects.

MICHAEL KARANICOLAS: Anything that you wanted to -- no?

UNKNOWN SPEAKER: (Off microphone)

MICHAEL KARANICOLAS: Sure.

JOHN LAPRISE: Hi, John Laprise, ALAC from NARALO.

I would just like to implore the group to reflect on the importance of UX and UI in the ODI initiatives. In everything we've been watching up on the screen today, there's been very little mention of UX in the design process. And this is really important if end users are going to be able to use the open data effectively.

Thank you.

JAY DALEY: So for the Open Data Initiative, this is -- this is about machine to machine, largely. This is about data scientists accessing and pulling the data. Many of those tools do include visualization tools or other ways of either understanding, expressing an insight or possibly telling a whole story through the data, but that's not a priority for this project. This is very much about the data scientists being able to access it.

And so the UX and the UI is very much targeted at data scientists and the way they're going to be able to pull that out specifically for the ODI project.

JANA JUGINOVIC: Thanks, John, for the question, and I know we had spoken at the At-Large session on Saturday.

The UX is really important, obviously, for the ITI project because it's the public information on -- that we're starting with on icann.org. And part of the reason why people have trouble finding information on icann.org is because of the UX.

We didn't do a good job of establishing those user journeys, how someone actually gets to the content that they need to quickly.

So UX is completely important to ITI. It's part of the project plan. And that's why, when we -- we'll be displaying that content and features on feedback.icann.org, we want you to not only comment on the functionality of the content but also the UX; you know, where you'd actually see a button, does it make it easier to access whatever you're accessing, whether it's public comment or anything like that. So absolutely important.

MICHAEL KARANICOLAS: Number 3, right in the back.

SATISH BABU: Thank you. My name is Satish Babu and I'm from At-Large from Asia-Pacific. We welcome the Open Data Initiative, and we think it's going to bring in a lot of transparency into ICANN's functioning.

I have a question on the license terms on which data is being released.

Do you have the details of what license? Is it an open document license or any of the creative comments license? Or what precise form of licensing do you have in mind when you release data sets?

Thank you.

JAY DALEY: So that is still to be decided but it will be an open license and so there are really only two proper candidates there, either Creative Commons or the separate open data license. The latter may be more preferable but there's still some work to determine which one.

MICHAEL KARANICOLAS: Number 2.

ISHAN MEHTA: Hi. Ishan from NextGen. I'm a newcomer at ICANN. So -- But I have worked on, you know, ICANN-related data before, and it sounds like there are a few other academics and business groups who have already, you know, built their own data sets peripheral to ICANN's work. And I would like to know if you guys have an initiative to try to integrate those data sets that are already existing by third parties or if not integrate, at least index them in your Open Data Initiative? Because I think right now, you know, our university hosts a small data set that might be useful to somebody else, but making that connection is harder to, you know, look up Georgia Tech ICANN rather than, you know, ICANN's Open Data Initiative. So that's one recommendation I have. Thanks.

JAY DALEY:

So we have looked at that, but no decision has been made on that. I have seen a number of open data platforms where that is done and is successful.

This is a complex environment, though, and there are some significant questions to ask about the authenticity of such data and whether it will be maintained and updated in other things.

And so I think there's, unfortunately, a long way to go before that particular question can be answered.

MICHAEL KARANICOLAS:

To 1.

FABRICIO VAYRA:

Thank you. Fabricio Vayra from Perkins Coie again. Thank you for letting me ask a follow-up question, and I appreciate the ITI explanation difference. ITHI. ITI and ITHI.

Maybe for the benefit of those in the room and since we're talking about the timeline of ODI, and I know we're going to get an ITHI session tomorrow, is there -- can we get a preview on what the timing is of the data that we've now been looking at for

two years of when that -- that will be available, where that is in the process?

What I'm hearing is the ITHI data is perfect for ODI and using ODI platform for publication. But just so everyone can see the two things in track, could we hear where we are with the ITHI data?

MATT LARSON: Sure. The ITHI project has been a long time in consulting the community and developing the metrics. So the metrics have just recently, I would say, almost solidified and we're just now starting to track -- track the data that the metrics specify, at least on the naming side. The number side, the RIRs are working there; perhaps a little behind in that schedule, behind where the names are. So at this point I would say we are dependent on the ODI platform to be significant with publishing ITHI metrics. So we're -- I would say we're, you know, months away, not weeks away.

MICHAEL KARANICOLAS: All right. We have time for one and -- oh, Raoul.

EN

RAOUL PLOMMER:	I would actually like to ask in general about the data management system. Is that going to be something that's free for everyone to use? Is it open source?
JANA JUGINOVIC:	It's a Document Management System, not a data management system just to correct. It's called Alfresco and, no, it won't be available to the community. It's a way of actually storing the content in this Document Management System, but the content is then surfaced through the Content Management System to the community.

The eventual path is once that infrastructure has been laid out for icann.org and those 15 other sites that are public information, is to roll that out for the SO/AC sites. So that Document Management System will be deployed and used for the content that's currently on the SO/AC sites.

MICHAEL KARANICOLAS: I see number 3 in the back.

CALEB OLUMUYIWA OGUNDELE: Yeah, my name is Caleb and I'm from Nigeria.

My question is the Document Management System and the Content Management System, I mean, the application (indiscernible) thing, are you going to make it open source so that all the developers can probably have a look at the code? And if there is going to be future development, we could look at it and make some contributions. And maybe we could use it for other data processes.

MICHAEL KARANICOLAS: That may have been -- It's very similar to the last question, but did you want to add anything?

UNKNOWN SPEAKER: No, that is not in the plans. For now the Document Management System is purely for the internal -- our internal content and making sure that it gets categorized appropriately so that it be in better position to be surfaced externally.

The same goes for the Content Management System. It's purely for the for the surfacing of content but no plans for it to be. Do you have additional?

DAVID CONRAD: Yeah, one clarification. Alfresco and the Content Management System are using .CMS. Both do have open source versions. The

version that we are going to be using is actually from Alfresco.com. It's a commercially licensed version that's internal only. The -- It is our data store, and we obviously would not want to make our data store available for external entities to go in and scribble all over.

So if someone is interested in the DMS or the CMS, the Content Management System, we're using well-known open -- well-known products that do have open source implementations that anyone can deploy if they so choose.

MICHAEL KARANICOLAS: So there are several ingredients to being a good moderator but the most important is never make people late for lunch. I can't see any other questions and we're just at the stroke of 12:00, so I'd like to thank all of my co-panelists for a very interesting discussion. Hopefully this has helped to advance the community feedback that we have all been looking for, and I look forward to watching these initiatives as they're taken forward.

[END OF TRANSCRIPTION]