
BARCELONA – ICANN GDD: Domain Name Marketplace Indicators

Sunday, October 21, 2018 – 09:00 to 10:15 CEST

ICANN63 | Barcelona, Spain

UNIDENTIFIED MALE: ICANN 63 Barcelona, 21 October 2018, 9:00 to 10:15. ICANN GDD – Domain Name Marketplace Indicators.

MUKESH CHULANI: For those of you in the cheap seats, you can move up front. We won't bite. I promise. We'll just wait for a couple more minutes and then we can start. Okay, we can get going.

Good morning, everyone. Thank you for giving up some of your Sunday for this session. My name is Mukesh Chulani and I'm the staff facilitator for the domain name marketplace indicators. With me is Amy Bivins who is also supporting with the project. And we have Matt Zuck and Hugh [Deaner] with us today to cover aspects of the project, which they're overseeing at this moment, to seek your input on specific areas, if you can advance to the next slide.

So, there's four things we wanted to go through today. First thing is to review our current progress against, of course, the timeline and our project scope. This is something I always start each of our meetings with. It's something that members of the advisory panel should have seen a long time.

The second one is to provide a snapshot of our current marketplace indicator schema. So, for those of you that don't know, we developed a

Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.

beta schema for marketplace indicators and we've been releasing every six months a beta report which showcases specific marketplace indicators.

Concurrently, we've been working with the advisory panel that we have to iterate that and make it a bit more representative of the specific factors that we seek to measure. So, you'll see there's been quite some change in the focus of the scheme and we'll give you a snapshot of that.

The third part of our agenda is really where we hope to spend the most time. So, we've developed now a draft version one schema. It's no longer going to be beta. And there's a few considerations that we need your input on and we will have Matt and Hugh discuss those. Then we'll go through our next steps, set you off on the rest of your day. Move to the next slide. That is not my slide. That's my slide.

So, project timeline. As you see, this is something which we've been at since the end of 2016. Since then, I've had the second boy. He started walking and he just started talking. We are now towards the end of 2018 and what we've done so far is we've put this beta metric out for public comment and then we've put – we've evolved the beta schema with a lot of support from the advisory panel, who have really been helpful. We talked through what the goals of this project were and the scope to make sure that everything we did was aligned to that. We started looking at specific metrics and definitions. We started to look at other considerations, such as the frequency with which we release this and what format it should be, whether we focus only on internal ICANN

sourced indicators or whether we go to third-party data providers as well.

Once we did all that, we wanted to get a third party to review what we've done, just to make sure that we essentially could have some solid footing. So, we got the support of Matt Zuck who is sitting here and will speak later. We're towards the end of this independent review, which Matt is conducting, and the next step will be essentially having the rubber hit the road, so to speak, so we actually start to look at data sources. If we need external data, providers will draft RFP if that's required and the goal is to release something, at least a wave one, within the first quarter of the next calendar year.

This next slide just kind of outlines what we've done for this independent review. It's taken quite a bit, so I wanted to break down what we've actually done.

So, we had a working session to evaluate the methodology of the entire assessment. Once we got clarity on what we intended to do and that was okay, Matt put his schema assessment report. We had a working session to evaluate that. Once that was okay, we started working on a taxonomy. A taxonomy document that we have essentially says, "Here's the indicator we want to measure. Here's the rationale for that. Here's how we calculate it. And here's potentially considerations for interpretation of those metrics."

We most recently conducted a due diligence of third-party vendors, so as we evaluated the schema, not all the indicators were going to be available internally from ICANN sources, so we started to look at third-

party vendors, conduct due diligence on whether the metrics we thought were available were really available and that's part of the discussion which Matt will go through today. We aim to finish this very soon, as you will see. Go to the next slide. That, again, was not my slide.

So, what's changed from the beta? I just wanted to provide a quick snapshot. This is very numeric kind of cold, hard presentation of what's changed. We've always had three categories for the metric schema. We want to look at what metrics can track robust competition, marketplace stability, and trust and we did not change that.

What has happened is we had nine factors underneath those categories, and factors are essentially statements underneath which you select metrics. So, we had nine factors and we kind of tightened that up to six. But we expanded the number of indicators from 28 to 45. We also realized that we perhaps ought to have some normalization variables and normalization variables essentially allow you to put a weight on your calculation so that you can adjust and compare regions a bit better. Out of those 45, 13 were carried over from the beta.

Forgetting the number side of it, what's really changed here is the beta was very much focused on contracted parties. It was very focused on the global spread of registries and registrars and there were some pitfalls to that. What you'll see is there's been quite a pivot to the registrant and also a wider focus of the entire ecosystem. So, we are not only looking at the contracted parties in this new schema, but we are also trying to see whether we can look upstream to the back-end service providers and downstream to the resellers.

We exclusively covered gTLDs in the beta schema. That was out of convenience because that's data we had internally and now we're going to include ccTLDs. Data sets, of course, as I mentioned, was sourced from within the ICANN organization in the data and now we're going to look at both within and external to the organization.

For those of you that have no clue what these schemas look like, you can access them. You'll see the notes there. I have a link to the data and the version one schema is in the appendix of this deck which you can get through the meetings page.

So, what's the new schema looking like? We retained specific metrics. Again, this is just illustrative. It's not exhaustive. There's 45 metrics. The font would be six if I put everything here. But we're covering. We're retaining the registration volume and change. We are retaining metrics that look at consolidation within contracted parties, accreditation volume and termination volume. We are looking at UDRP and URS volume, complaint volume and the decisions. That's retained. And incremental to that, as you'll see, there's a broader scope here, so we want to look at registrant geography, language diversity. We want to look at service provider language availability and payment method diversity. We want to look at renewal activity, churn. We want to look at use category and we have this idea of perhaps we can track adoption as well. Matt will cover that later on. That's one of the items we need your feedback on. And a few other things you'll see there. I don't want to go through everything on the slide. I want to leave Matt enough time to go through his section. You can go to the next slide, Amy.

That's my brief introduction. We can save questions until the end, if that's okay for you. I will turn it over to Matt to go through the taxonomy report and you can take it away.

MATT ZUCK:

Just to confirm, it's 10:15 that we go to? Okay. Just want to make sure I had the timing right and can pace this accordingly. Anyway, it's great to see so many people here today. I want to thank everyone who has been participating in this beforehand. I had some nice conversations with people at the Panama meetings, as well as by phone and e-mail conversations. I really have focused on taking what people have brought as the advisory panel to this overall taxonomy, these particular metrics, and sincerely hope that you all can see your hard work reflected in this overall set of taxonomies and set of metrics and so forth.

As Mukesh was saying, we're not going to have time within this particular meeting or this particular session right now to go over all the 45 metrics. I'm going to be focusing on some of the outstanding concerns that have come out, particularly as we started moving to looking at the vendor providers for this, what's available in the marketplace for some of these things.

I will say that I am planning on being around after this particular session, so if people want to buttonhole me right after that and have a longer conversation than we're able to have right now, I'm very happy to do so, and of course by e-mail and other means as well. I'm trying to

think. I think that's pretty straightforward, so we can move on to the next slide.

Okay. As Mukesh was saying, one of the things that was part of the current phase we're in here is starting to look more seriously at the source of the data for these indicators. Lots of interesting potential for indicators. Lots of interesting metrics out there, but in the end, we still have to source it from some place.

The inclination has always been to go with data that is internal to ICANN for all kinds of reasons. Consistency, availability, cost, and things like that. So, for those metrics where it's available, it's about 29 in total, we're going with sources from within ICANN.

The three normalization variables are available from pretty standard external sources freely. I think World Bank sources for those. That leaves us with remaining 16 metrics that draw at least partially, some completely, from external data vendors.

The past two months since we had the last phone call, or the conference call, the advisory panel at the end of July, we've been going through an evaluation of looking at different vendor providers or in the marketplace looking at different vendors who can provide some of the necessary data for this. It's actually been quite interesting to see what is available, how it conforms to the vision that we have for this particular set of metrics, but then also a lot of the sort of outstanding questions that are here. I think this is both – it reflects the creativity of the advisory panel, thinking through what would be good indicators, what should be part of this schema. But, also, the fact that there are

some decisions that have to be made. There's data available for all of these categories, but depending on how things are defined, you start getting into some differences that can be quite significant and I'll get into that in terms of what the data looks like, what you end up reporting.

So, this is where, in particular, I'm very interested in the kind of input from the people who have come here today, both the formal advisory panel but everyone attending as well in terms of these sort of decisions, how we might go about doing this and that's what I'm hoping to take from this kind of conversation and get your feedback and also reflect the intent and the hard work that's gone into this so far.

Just a quick highlight. There's going to be four areas of discussion that I'd like to go over today and we'll move through these. The first one is probably the one that has the most significant discussion worth having. This is in the topic of resellers. It impacts a number of variables in terms of the questions we have today, has the highest number of variables that are impacted directly. You can see them at the bottom of this slide what the definitions – not the definitions, but the names of these particular variables. And essentially, it gets to topics on the number of resellers by ICANN region, the percentage of all gTLDs that are being registered by resellers as well as the marketplace concentration information as well and some other things.

So, this is a really important point. This is something that was talked about quite a bit by the advisory panel in terms of something that wanted to be included in this set of indicators. It's also somewhat

difficult because tracking resellers is not within ICANN's current remit, so we get into a definitional question. And definitional questions are fine. They're very straightforward, but in the end, we have to think about what kinds of – how exactly do we want to define this particular category, which is the conversation, the input I'm very interested in hearing from people in this room and as well as accessing remotely on this.

To summarize this based on our review of vendors who are providing this kind of data, there are two main approaches, and I should say that there's certainly some nuance within these different approaches, but for the sake of this meeting, I wanted to divide it down into these two different approaches on how we go about measuring resellers.

The first is relying within WHOIS data records that has a particular category for resellers that identifies a particular domain having been sold or being associated with a particular reseller. That would be quite nice for aggregating it up and getting to the overall percentage of gTLD domains that are being run by resellers or being sold by or registered by resellers.

Now, that's a great plus, it's a great advantage because it's a field that can be populated or not. The issue is that there's a relatively small number of domains for which this field is actually populated. This is of concern because the decision to populate this or not is at the level of the registrar, so this is not something that is necessarily used exclusively across gTLDs and across registrars. So, while you have a sample of this sort of reseller, you certainly could use this to look at

number of resellers by ICANN region. It is problematic that this field is not consistently used across registrars, so you end up with sort of a bias sample on this. I think this would be ... So, that's the positive and the negative of this particular way of approaching for resellers. You do have this nice way of counting it, but there's a bias in the resulting sample from it.

The second one in terms of defining resellers is looking at hosting records and name servers for domain. There's various ways of doing this, but essentially it's looking at the name server records to identify the clusters of domain holdings. This captures the full population of gTLDs because it's based on this sort of technical definition of name servers, but it a more troubling way to disambiguate what these clusters actually mean. Is this a particular cluster? Is this capturing ... I guess another way of saying it is what do these sort of clusters of domain holdings represent? Is this simply something that comes about because of reseller activity? But there's any number of other kind of activities that we might see as well in terms of small domain or web development shops or larger web development shops, as well as some sort of brand protection strategy as well for this.

So, the current data we have here, the way that one could disambiguate or say that this group of domain holdings are resellers versus not is largely through a numerical basis. You could come up with some sort of threshold, be it 50 domains within domain holdings or 100 domains and say that this is a category that, above this number, we're going to classify as resellers. Below this number, we're going to classify as something else.

The inherent problem in that, of course, is that simply by using that sort of numeric measure, you're going to capture a lot of things that aren't necessarily resellers with that particular definition.

So, these are ... Having gone out to look at field, this is what we have in terms of resellers. Again, I think the advisory panel is quite right in saying that resellers are an important part of the domain name market is something that we'd want to be talking about and looking into, but these are the categories we have here. Should I go through and see all these and open it up at that point?

MUKESH CHULANI: We can open it up.

MATT ZUCK: Yeah. Okay. I would say, just for this introduction, because I think this is also one of the points that's probably the most interesting discussion, to sort of open it up to see what kind of input and feedback people might have.

UNIDENTIFIED MALE: Thanks, Matt. Go ahead, Mukesh.

MUKESH CHULANI: So, I think if I just want to clarify, our situation here is we know we've gotten the feedback that we should track downstream for contracted party to see whether there's stability or robust competition and trust

within that reseller space, and as per due diligence, we found two ways to do it.

The first one is to go through WHOIS. Not all registrars fill that space in. Some of them do. So, that's one opinion. The second opinion is we look at hosting records and name servers. We could also do that, but the caveat there is we would have to define what the reseller is ourselves. With WHOIS, it's explicitly stated, but not everybody tracks it. If we go to opinion two of hosting records and name servers, we can cover the full population, but we are the ones who are making up that definition. So, that's the kind of situation we're in. We open up to your thoughts.

ROLAND LAPLANTE:

I'm Roland LaPlante with Afilias. Sorry I'm not familiar with what the advisory council has advised here, but what surprised me here is now we're talking about we don't have and I thought we were going to talk about data we do have. We have a pile of data. We have all the registry operator reports going back a decade every month. We have all the accredited registrar reports that go back forever. So, I don't know why we can't come up with some indicators that are based on the data that we already have. There's plenty of issues in that data, believe me. But, I wouldn't hold any of this up for reseller data which we do not have. I mean, for me, this is a boiling ocean kind of thing. There's thousands and thousands of them, hundreds of thousands probably of resellers, and they change every month. It's seeking the grail to get into this element. I mean, I don't dispute that resellers are an important part of the ecosystem, but I think as we start to have marketplace indicators

publication regularly, that we should be looking at data we have, data that we know about, data that we have some sense of whether it's accurate and timely and complete and deal with what we already have because it's free. We already own it.

MUKESH CHULAI:

Thank you, Roland. This is very, very useful input. Unfortunately, the rest of the deck will be covering stuff we don't have. That's by nature. So, there's 45 metrics and we are left – of those 45, so many of them are internal to ICANN. The exercise here was to look at the subset which ICANN didn't have and then see who was able to provide that data and that's the specific items we're going to cover. There's four themes I think we're going to cover. There's four themes. Reseller is one of them where we went out into the marketplace to look at third-party vendors to see what they could offer in this space. That being said, your feedback is very, very welcome and we thank you for that.

UNIDENTIFIED MALE:

Also, I might just say that some of the data you were just referencing right now actually is within this report and it is going to be put forward on—

UNIDENTIFIED MALE:

It's like 30 of the indicators are internal data, right? 31 or something.

UNIDENTIFIED MALE:

Yeah, it's 29.

ROLAND LAPLANTE: Okay. So, why don't we just start on that? I mean, if we focus on the external stuff, no doubt it will eventually be important and helpful and insightful and so forth, but we may never get to the stuff that we actually have today if we focus all of our activity on stuff we don't have. Anyway, I come to these meetings and I always make negative comments, and I want to tell you, I'm fully supportive of this overall effort. I'm just trying to guide it in a way that will get us something useful sooner rather than later.

MUKESH CHULANI: It's noted. Thank you. Thank you, Roland. I think we have a comment as well. Amy?

AMY BIVINS: Yeah. We have a couple of comments in the chat from John McCormick. The first is that there are about 1.4 million [inaudible] resellers. There's only a few thousand registrars. But not all resellers are registrars and a lot of those are domain hosters.

MUKESH CHULANI: Is there any other feedback? Of course, Ashley.

ASHLEY: Hi, Ashley here from [inaudible]. I just wanted to say that even number one, requiring resellers to fill in the WHOIS field is, first of all, not

something that I would want pushed. Second of all, it's not accurate. Just because we enter one reseller into the field doesn't mean that they don't have resellers or their customer doesn't have resellers, so there's no way to try to track this whole chain even using field number two, trying to find it through the data. So, I would support the previous comment. Definitely don't try to go down this route. Try to use the data that we have to understand more and try to clean it up rather than trying to find this little needle in a haystack.

MUKESH CHULANI: Thank you, Ashely. Matt, you have—

MATT ZUCK: I think this is the kind of input we're trying to get back from this because, again, we went out and we saw what was available because some of the previous sessions have been ... To stress this idea that we want to have resellers. This idea that resellers should be part of this, so we went out with this trying to see what could we get. If we wanted to bring this data in, what was available? So, that's how this is being presented today and wanting just to get this kind of feedback from people on whether or not ... What issues? We know some of the issues, but that's a perspective that's quite useful as well.

MUKESH CHULANI: Alright. Let's go onto the next one. And I think just before I get into this, I will say that, again, to this point about the data that we have that we don't ... Again, we are focusing on the data we don't have right now

because we're looking for this input, but I would be more than happy to go and – not right now, but talk about what we're doing with the ICANN data that we do have, the internal ICANN data, and that's actually at the end of this slide deck. So, there is a lot of work there. I don't want to ... Not focusing on that today because of this perspective, but I do want to acknowledge that it's there and I think there's a lot of interesting stuff that's going to come out of this.

MUKESH CHULANI:

So, again, that's why we're putting it out to the advisory panel because we're stuck on these items. The items we're not stuck on, we don't need to discuss. We've all gotten consensus that it's good, that it's useful, and we have it. So, of course, we're going to go with that. These are the last few items which are sticking points.

MATT ZUCK:

So, moving on, the second of the four areas we wanted to talk about. This is impacting a single metric. This is RFC 2.7. Essentially, to give folks some history of where this particular metric came from, since I suspect there's people in the room who haven't been in this from the beginning, this is really from the advisory panel's perspective wanting to get at the final use of a domain, the end user of domains. How are domains actually being put to work after they're registered, purchased, and hosted? What's actually being done on these particular domains?

There was considerable discussion about how this should be classified, what kind of sets of classifications. A lot of discussion about the issue of

including a definition for parking or not parking. I'll come back to that in a little bit.

But, the way that we have framed this particular metric in terms of looking at domains that are resolving, not resolving, resolving with the same TLD or being redirected to another TLD, the intent was for this to get some well-specified measures of domains that would be possible to collect, not dependent on more specific measures or definitions, but essentially that if we went to any vendors collecting these kind of data that we'd get the same sort of results. They're very specific definitions that are less based on any particular vendor's idea of what parking is or what for sale domain names are. Again, I'll talk about that in a bit.

Again, the idea was ... So, the basic measure is a domain resolving, not resolving. Then, is a domain resolving within a TLD or is going, being redirected to another TLD? The intent behind that was, one, it's easily measured and the second, it was possible to see are certain kinds of domains more likely to be set to a home domain. Are these defensive, branding kind of opportunities that are happening to get at these sort of measures?

Now, that said, we went out and we found vendors where we could get – the way it's being defined now fairly easily, fairly in a standard way across these definitions, but there are certainly more nuanced categories out there. And just to name one that has generated quite a bit of discussion earlier within this group is parking.

Just to think about the ways people might go about ... And just to give you a sense of some of the different definitions we've found within

vendors on this, you might see people distinguish between parked domains and domains that are listed as coming soon or domains that are being parked domains, domains that are coming soon, and domains that are listing for sale.

Again, this gets at the endless creativity of human beings, on what they might want to do with a domain, but also gets to the definitional question of how do you distinguish between a parked domain and a domain that might be being listed for sale? You certainly could do this and people are certainly doing this. It's possible to pick a vendor and say, "We're going to go with this vendor's definition of this." But again, since people are making different decisions as they're creating these statistics and these data measures, you're going to end up with ... We end up with different numbers in terms of the number of domains parked or the number of other kinds of ways we might want to specify how domains are being used.

So, again, this is something we'd seek some additional input from you all as the advisory panel on how we might push beyond the way it's being defined right now, to some more nuanced categories to the extent to which that's possible or not. So, I think I'll turn things over for comments on this particular point.

MUKESH CHULANI:

Matt, could you clarify what your actual question is?

MATT ZUCK:

Sorry. Always clear in my mind, not necessarily verbalized. What I'd be seeking from this is the discussion on the extent to which people would help – essentially looking for further specification from the advisory panel on some of these nuanced categories. How should we go about defining what parked is, for example? But not necessarily limited to that question.

ASHLEY:

Hi, Ashley again from [inaudible]. I don't know if you're aware of this, but there's actually a ccTLD group that's working on creating standardized metrics across various ccTLD registries, so define what a renewal is and what a delete is, what a restore is, and coming up with industry-wide definitions that will be used by ccTLD registries. This is also a question that they've taken a look at and I think they also have been working for months on creating more standardized definitions of parked page, low-content page, and a number of other redirect issues. I can put you in touch with somebody from that group and they can share what they've been working. It doesn't hurt to, as a registrar, to [look at] multiple data sources and be able to compare them because the definitions are the same. So, I don't think anybody would be offended trying to share that standardization across the gTLD world as well.

MUKESH CHULANI:

Thank you, Ashley. We are actually plugged into that work which Patrick, Miles, and Neal McPherson are doing. We are going to chat with them as well. So, that's going to happen. I think our situation now is we

have ... These are measures intended at evaluating usage and adoption of gTLDs – sorry, not just gTLDs, TLDs. So, it would cover ccTLDs as well. I think the first part is it's kind of possible to do this resolving, not resolving. So, are the lights on are not? Then, redirecting within or outside, so that's okay.

There are some sub-categories which we don't know whether you think would be valuable to track as well. So, should we also start to look at parking or not or shall we just leave that aside and focus on what's already feasible at this point? But, thank you. Thanks for the feedback there.

AMY BIVINS:

We have a comment in the chat from John McCormick. "The ccTLD group isn't up to speed on the parked page, low content page definitions. It's too early to standardize."

MUKESH CHULANI:

So, John, it's Mukesh here. Do you think we should stay out of parking at this point then? Is that what you think we ought to do? That's not just a question for John. That's a question for everyone here, including the cheap seats.

AMY BIVINS:

John responded in the chat that, at the moment, there's not a single definition I think of parking. And he's typing, so he may have more.

UNIDENTIFIED MALE: Hi, it's [Deepmat] from [inaudible]. Sorry, I might be missing something, but we're only looking at whether a domain is showing a webpage, is that correct? Or are we looking at things like e-mail, for example, and other types of usages that don't show up from a webpage?

MATT ZUCK: The way that we've been looking at this, the way this is defined right now, is just looking at the webpage definition. There's certainly an argument to expanding out in the way you're talking about as well. Would you be promoting that?

UNIDENTIFIED MALE: Yeah. I think that's actually ... It might be a bit more worthwhile because you might be doing a test and come across pages that are blank, and for all intents and purposes, you [inaudible] actual usage for that particular name, that second-level name. But in actual fact, there is actually usage because the person's made a conscious decision not to put up a web page and only use it for mail and that conscious decision isn't necessarily a bad actor. It could also be a normal human being on the street that's just decided, "I don't need a website. I just want to use it for mail." So, I think that is pretty important.

MUKESH CHULANI: Thank you. That's pretty helpful feedback. Going, going, go ahead.

MATT ZUCK:

Alright. Moving on to the next slide, this is impacting again a single metric. This is impacting T1.7 which is looking at the percentage of domains using a – sorry, let me [inaudible] so I can read this. Percentage of domains choosing privacy or proxy WHOIS services by gTLD category. This is something that before we started looking at vendors, I was more hopeful that it would be possible to come up with a fairly standardized definition of it, but part of the issue with this is that, again, since there's not a standard way to implement a privacy or proxy WHOIS, and particularly this is getting more complicated with the recent legislation in the EU in terms of privacy, that there's not ...

Again, there's not one way of measuring this or one way of defining this and having talked to some of the vendors and having tried to do this myself in various ways, it is quite challenging in terms of how you might go about doing this, in terms of examining WHOIS records. You can certainly do things like looking for key words such as proxy or privacy. That's not necessarily always used. Other ways of doing this is to aggregate up on the actual registrant name.

If a certain name or text string is associated with lots of domains, it's either an indication that there's someone who's registering a large number of domains or some sort of privacy and proxy thing. You would also do the same thing on e-mail addresses that they might be associated with because a large number of domains associated with a particular e-mail address, that's worth taking a closer look at. So, in some ways, this approach is looking for outliers and taking a closer look at the data associated with those outliers.

Of course, doing this, you can approach this in all kinds of different ways and different people of good intent can come to different conclusions on what should be included in this category and not.

So, as a result, the numbers that we've been able to look at and looking, comparing across some of these data sets, the numbers will differ between the data sources.

Again, this is not because they're necessarily even trying to define things. They're trying to get at this measure of privacy and proxy, but they're just making different criteria by which a domain would be categorized this way.

So, it becomes sort of an issue of ... Again, it's not an order of magnitude we're talking about here, but we would end up with different numbers when we go towards this.

So, the question I'd like to turn over to this group, the advisory panel, is thinking about is there a standard way of measuring this privacy-proxy WHOIS that people are aware of and would promote for us to move forward on this that we could ask vendors to apply for or apply to their data generation?

UNIDENTIFIED MALE:

Can I ask you to clarify? So, what you're saying is different vendors draw different lines in the sand for a definition of privacy-proxy, can you give some examples of what those lines in the sand are?

MATT ZUCK:

I mean, the different lines in the sands. Again, there could be – an example of our WHOIS record that the registrant information might differ, so if you were just looking for differences by the name of the registrant, it might not show up, but if you were looking at the contact information for the registrant, the owner of the domain, you would find an e-mail that’s not actually the e-mail of the registrant, the person who registered it, but it goes to standard e-mail within a registrar.

Now, some people might classify that as a privacy-proxy. Other people might not be as attentive to that and not include that. So, you’d end up with different classifications on domains.

ROLAND LAPLANTE:

Matt, is anybody looking at the ... You mentioned the GDPR and I know at Afilias ... Sorry, this is Roland again from Afilias. I know at Afilias we stopped publishing a full WHOIS back on the 25th of May and now you couldn’t tell anyway if you looked at our WHOIS. So, you have to get a private feed in order to do this analysis, and in order to get a private feed, you need some rationale that says you have some legitimate reason to have personally identifiable information. The security community is having trouble coming up with this rationale and I think this is less important than the security community use of it. Is anybody looking at this from here? Because this may actually not be possible now.

MATT ZUCK: Yeah. That’s a good point as well. The conversation we have with data vendors were suggesting that they were able to work with this. Now, whether the extent to which they were overly hopeful or overly hopeful or overly optimistic about this I think is ...

ROLAND LAPLANTE: Can they work with no data? I mean, that’s essentially what they’re going to end up with.

MATT ZUCK: Yeah. And this was one of the things that came out of the conversations as well. The GDPR, that was making this complicated. This was something that was put in before those issues came out. So, I think it’s a legitimate question and I think it’s something that we’ve not resolved with the conversations [with] data vendors.

ROLAND LAPLANTE: And does this get at the trust category? I’m trying to figure out, is this competition or stability or—

MATT ZUCK: This is trust.

ROLAND LAPLANTE: Okay, thanks.

MATT ZUCK: One of the reasons we kept this in, too, because we think it's ... We wanted to keep some of these trust measures in, too, to the best extent that we can. But I think this becoming increasingly problematic.

MUKESH CHULANI: I think Amy can give us some input as well because she's working on the privacy-proxy effort.

AMY BIVINS: Yeah. This is Amy Bivins from staff. Just so you know, it's not available now, but there's a proposal to label all privacy-proxy registrations consistently once we implement the privacy and proxy accreditation program. So, that wouldn't solve it now, but once it's implemented it could do that.

We also have a comment in the chat from John McCormick. "The GDPR mess has created problems with working with WHOIS samples. [inaudible] domains are now under privacy I think."

MUKESH CHULANI: Please go ahead.

UNIDENTIFIED MALE: Hi, it's [Deetmar] again from [inaudible]. I'm not making a judgment as to whether privacy-proxy is good or bad, but what's the expected outcome from looking at whether somebody has privacy-proxy for their WHOIS or not? Because some individuals may go through a registrar

and just choose the box that says “hide my data” and that might fall under your definition of privacy-proxy. So, what’s the actual plan at the end of this? Are you looking to see whether they are only bad actors using privacy-proxy or what’s the actual outcome?

MATT ZUCK:

The rationale is a bit more broadly than that in terms of thinking about trust in the marketplace. I’m just going back to looking at the rationale for this. Basically, it’s providing information, the ability for the community to access WHOIS data and actually be able to contact the owner of the domain, sort of what WHOIS was originally for. The extent to which that’s actually possible to do within what’s happening and that was sort of the intent. That was the rationale behind that.

It wasn’t really trying to get at bad actors or not. It was really just the idea that the important part of trust within the marketplace is, well, if you run into a problem, can you actually get to the person who is associated with the problem?

MUKESH CHULANI:

And just to add why we’re even looking at this as an indicator is because of the category definition we’ve chosen for trust. So, trust, as we’ve defined it, is that the domain industry demonstrates operational success in safeguarding community interests, including registrants, intellectual property holders, and law enforcement. So, that’s why this indicator is even there.

MATT ZUCK:

Okay. Moving on to the next slide, this is the fourth of the issues we were wanting to bring up and get feedback on. This one is tied to metric RC2.8 which actually is sort of an expansion of the metric I was talking about two slides back, metric RC2.7 which again the intent, as it came out of the advisory panel, was trying to understand end user behavior. How were domains actually being put to work?

So, we had the one set of metric that we split out and were looking at the very technical definition of resolution or not resolution of being forwarded or not, and this one was trying to capture some of the more nuanced ways that we might go about doing this [inaudible] specified is a TLD specific strategy for branding. I'll talk more about what we mean by that, but one of the things as we were doing the review for data vendors, we put this out and this is probably the one that, the metric that we were hoping to get the most impact or input from vendors and see what was available because we weren't quite sure the extent to which this is being tracked or not by different data vendors out there.

We quickly ran into some pretty obvious problems, which we knew going in. But just the issue of how exactly are businesses to be defined? There's businesses that are sole proprietorships all the way up to large corporations and there's certainly various projects that are the dividing line between personal project and actually active business can be a very fuzzy line. There's differences across region and country in terms of what gets defined as a business.

Then, there's also the issue of how a TLD-specific strategy was operationalized. In putting this together, we were thinking primarily as

an example, but not necessarily limited to a decision by a firm to pursue both a generic and a country-specific kind of solution, have a sort of localization through a ccTLD and where they're operating but also using a more generic or gTLD for their overall branding and sort of strategy, though it certainly wouldn't be limited to that as well.

Now, what we found from this – and again, just to give you a little bit more background, one of the approach we were thinking about was going with some sort of omnibus survey which would certainly solve the problem in terms of the definition of businesses and that would be turned over to the survey provider. We had more concerns about the extent to which the sampling that we'd be doing in this time would actually get to many firms that were thinking about a TLD-specific strategy.

So, in talking to vendors, we came across some various interesting data sources that essentially are trying to cluster domains by ownership and associations. This is different than clustering domains by name server locations or hosting locations. This is more looking at a single enterprise and which domains do they control. You can certainly sync this in terms of branding [inaudible] or larger companies but also in terms of some smaller company approaches to this as well.

The problem with this is that, of course, that while it's quite useful if you're trying to find out which domains, what domains a particular company or enterprise controls and how they're trying to market themselves, it's a bit orthogonal to what this is trying to do which is trying to rather look at the enterprise is trying to look at the market.

So, one could see how you could build from this collection in saying looking at how enterprises, which enterprises or which domains are owned by which enterprises and then turn it 90 degrees or so and start looking at how domains are being used across the market. That's not necessarily an easy step and there's no ready data sets available to look at this particular question.

So, the question I would ask for the people assembled here in the advisory panel is what other ways might we get at trying to get at this question again, at the end use behavior of domain owners and look specifically at how they're using these across TLDs. Again, with the previous metric 2.7, we can get at to the actual behavior and seeing how domains are resolving or being redirected, but this is also trying to see how we might – again, how we might get at this question. I think a very interesting question of how people are actually using domains and whether or not people have specific TLD strategies.

MUKESH CHULANI: Steve, please go ahead.

[STEVE MARR]: Hi, it's [Steve Marr] from [inaudible]. Apologies, I'm not going to answer your question. I actually have a question. If you're collecting all of this data, you're getting an idea of how a corporate or a business might register names and what types of names they might register. That might not be something that a corporate or a business would like the rest of the world to understand what their naming structure is.

So, I guess the question is the data that you're collecting, that's staying within ICANN, within a database, or is that actual physical data, the full data set, going to be made public? So, in other words ... I can't even use XYZ anymore because that's a real thing. Random company has a naming structure and they use random company and then RMD and RMDMCPY and you're able to collate all of that information into one group and you're able to identify that that's one single company, is that data going to be made public? I know it is public in a way at the moment because if you spent a lot of time and built an algorithm, you could probably get all that information, but in this sense, you're kind of making that information public quite easily, if it is public.

MATT ZUCK:

The easy answer is, no, it's not going to be available that way at the country level, and even further, what this is going ... This is more a description of how the data is now available because this is obviously a data product, as you say, that could be quite interesting to people. This is sort of a product that vendors would have put together and actually, as I was saying, it's actually 90 degrees off or maybe 180 degrees off – I'm not quite sure what the rotation would be – from what this actually would be because we're not actually interested for this in what a single enterprise are doing. If you could think of it, if we got this information by all the enterprises, aggregated it up, and then looked across the marketplace, that's what we would want to do and that's what this ... But yeah.

I'm not quite sure how feasible that is because there's lots of contingencies, but that would be the intent. So, yeah, the simple answer is no. That would not be available at the [inaudible] for a single enterprise and that's not the intent.

MUKESH CHULANI:

Just to provide some context to why we're even looking at this is our category definition, one of the category definitions for robust competition is that registrants are adopting domains across all TLDs. So, we looked at adoption and you saw a previous slide where we looked adoption as lights on, lights off staying within this space or redirecting to another space. That's one way we looked at. And this is the second way we thought or at least we received that input. This is another way through which we could gather adoption is what subset of businesses are actually branding themselves through the TLD. Of course, businesses is so broad. How do you even gather that? it's a very, very broad definition right now. That's why we're asking these questions. I think we have a comment.

AMY BIVINS:

Yeah. We have a comment in the chat from John McCormick. "There was an ex-Google guy who set up a site to map cluster ownership a few years ago. Can't remember the name or his name but will try to dig it up."

MUKESH CHULANI: Thank you, John. That would be helpful as well. Any other feedback on this? You need to stretch? We're almost over. Thank you.

MATT ZUCK: That was the ... This is going back to Mukesh's, turning things back to Mukesh, I guess I just want to say that what I was presenting here today is stuff that questions at – in some ways, people call it, sometimes referred to as wicked problems. They're the ones that don't have necessarily one solution or something that you can resolve and why I wanted to bring these up and get the feedback and some quite useful feedback we've gotten so far today.

But, in terms of if people want to talk with me afterwards about these questions or some of the indicators or metrics we have developed, I'd be more than happy to and I'll be around after this meeting, as well, but I'll turn things over to Mukesh now.

MUKESH CHULANI: Thank you, Matt. So, our next steps are to essentially capture all the inputs we received and you have some additional time. Matt is here and I'm here as well to capture any thoughts you might have. We hope to lock the taxonomy document down by the end of October, so we'll circulate that to the advisory panel. I think we've got some solid input in this session.

Then, as I mentioned, the next step is to actually review the outcome of this project internally within ICANN. So, we're going to look at how we

take this forward and obtain [inaudible] to actually start gathering the data.

As I mentioned, our plan is to release a wave one, so this is something which Roland I think also spoke to. We can't release everything all at once and we shouldn't wait until we have everything. So, we will start to do this in waves as we start to collect the data. Wave one we envision to have that within the first quarter of the next calendar year.

So, with that, I will give you all 15 minutes of your life back, unless you have further questions, comments. Of course, Roland.

ROLAND LAPLANTE:

On the internal data we do have, I spent a fair amount of time with the registry operator reports because I'm trying to get a sense of what things are happening, particularly what's going on with DNS query counts and I can tell you that there are TLDs that don't report at all, TLDs that report whenever they get around to it, TLDs that report what is clearly bogus information. I mean, it's just not even realistic at all. I know ICANN was talking earlier about maybe hiring a data scientist or having some kind of a team that would go through the data that exists and either try to clean it up or get whoever is reporting it to go back and report it accurately because in order to get any useful information out of this stuff it takes a lot of time to clean it up and if you don't take the time to clean it up, if you don't look at it closely enough to know that it needs to be fixed, you're going to make wrong conclusions.

And as we get more data-driven in this community, the accuracy and completeness of the data becomes really important, because otherwise, the community itself will make bad decisions policy-wise and other kinds of things.

So, is there any progress being at ICANN to resource the looking at this data and the fixing of it?

MUKESH CHULANI:

So, I think the first statement I want to make is we are not envisioning to release these metrics on such a frequent basis that the correction of such aberrations is not possible. So, [inaudible] looking backward and we will do it potentially twice a year or four times a year, so it's not like we'd release data on a daily basis and ... The risk of aberrations not being caught is much less because we're looking backward with such a lag. That's the first statement.

The second statement is I think once we start to actually churn these indicators we will start to notice what's funky and what's not. We are not doing this calculation now. I don't know what the numbers will come out as. Part of any release of these metrics will be an evaluation of whether we're confident with what these numbers turn out to be. If we don't have that sense of confidence, as you say, there's a risk of misleading, misdirecting. So, you can be assured that if there's funky data that we would examine that and we would put that to the advisory panel again to decide whether we should go ahead with it or not. At this point, our next stage is to commence this.

I think we have enough feedback from you now. We have a solid basket of indicators. So many of them are internal to ICANN. We should start looking at them. And then we'll continue to call on the advisory panel for support if we come across funky stuff.

UNIDENTIFIED MALE:

With the current state of the data, if you just report it the way it is, there will be incorrect conclusions drawn. If you clean it up, you risk people not trusting ICANN because they will think that you cleaned up the stuff that doesn't agree with your point of view. So, there's kind of a mess here and just having accurate, complete data would resolve that because I think we all want the raw data. We don't want data that's been massaged and otherwise made not raw. So, I think getting an accurate data set is really essential for the trust of this and a whole bunch of other efforts.

MUKESH CHULANI:

Thank you, Ronald. I think you can be assured that we wouldn't bulldoze our way through any of these indicators if they looked funky. There was one more thing I wanted to say, but I forgot. There is a comment from John McCormick to save me.

AMY BIVINS:

Okay. So, the comment from John McCormick is, "Registry reports don't always match DNS data. A lot of domain sites are not set up in DNS and it varies by TLD."

MUKESH CHULANI:

Okay. I remember now. The idea of the wave is also that we ... It's not always about sample of convenience. It's also about this reliability thing and that's why we would go with waves. I hope that's enough of a measure at least for us to get some. Thank you, all. Thank you for your time. And I think we can close the session now.

[END OF TRANSCRIPTION]