Transcription ICANN Los Angeles
IDN Variants Update to the Council
Saturday 11 October 2014

Note: The following is the output of transcribing from an audio. Although the transcription is largely accurate, in some cases it is incomplete or inaccurate due to inaudible passages or transcription errors. It is posted as an aid to understanding the proceedings at the meeting, but should not be treated as an authoritative record.

On page: http://gnso.icann.org/en/calendar/#oct

The recordings and transcriptions of the calls are posted on the GNSO Master Calendar page

Jonathan Robinson: All right, let's draw a line under that section now. And if we could stop the recording on that and prepare for the next session.

Can you give me an indication when we are ready to go with the next from a recording point of view please? Thank you. So we're good to go with the next session.

You'll recall that this is about IDN variants, and this is the work that's going on and an agreement that we came to after a couple of discussions about keeping this at an appropriately high level on the agenda and we agreed that there would be regular briefings as to what work has been going on in the IDN variant program.

And so we here we have the opportunity to have that briefing for the Council and the community. So I'm not sure who's - okay, hi. Go ahead then that's Sarmad is it?

Sarmad Hussein: Thank you. So what I'm going to do is take you through a few agenda items. We'll start with a general overview of what the IDN program is. And then we are doing quite a few things but I will, for today's presentation, focus on the update on the IDN TLD program and their share somehow reach efforts which we've been doing to the community, and then finally concludes with what plans are going forward.

Next slide please. So as far as the IDN program is concerned, generally we are working in multiple areas. One of the main thrust of our work is on what

we call the IDN TLD program which I will explain in a little more detail during this presentation.

And the program obviously has focused on generally trying to figure out what a valid top level domain label is and whether it has variants are not for various scripts.

We also assessed in that IDN ccTLD stream evaluation process in the IDN program, work on IDN implementation guidelines. And finally we focused on outreach to the community. So these are some of the tasks which we undertake with the IDN program. Next slide please.

So focusing more on the IDN TLD program, basically before the IDNs obviously came around at the top level we had ASCII domain names. And even when we had just ASCII domain names they were interestingly with different roles for domain labels, at second level was as top-level.

So for example at a second level the policies were a bit more liberal wherefore example hyphens and digits were allowed in addition to letters. But for the top level domains even at the outset the rules have been much more conservative meaning that even for ASCII domain names at top level only letters were allowed; hyphens and digits were not allowed

Next slide please. Going into now other scripts and we're talking about top level domains, this becomes a challenge because when we take the traditional restriction or constraints for the top-level domain which is restricting top-level domains to letters, we need to then, when we go to other languages and other scripts we need to have a very clear definition of what letters are in those scripts. And it's not always as straightforward as the ASCII quote.

So and then once you define what the letters are in the ways these scripts are encoded in Unicode, sometimes some of these letters are visually confusing to other letters so you not only need to define which letters are

eligible for top-level but which of those letters may be confusing for users and therefore should be called variants of each other.

And then finally we also need to see whether there are other additional roles which need to constrain. So for example, in some scripts they are combining marks and obviously a label at top-level should not start with a combining mark, it must start with a regular letter.

So some of these additional constraints need to be figured out as well, and these need to be figured out for each and every script which we used to define top-level domains. Next slide please.

So there was a whole program which was instituted to look at what the challenges are and how to solve those challenges. And it started back in 2011 where six case studies were done to really try to understand what the problems were.

These case studies identified the issues which were integrated into a community of report. And based on those issues there was another series of projects which looked at how to possibly solve those issues.

And out of that third phase came out what is called an LGR development process. Next slide please. And what was decided was that there should be a central - in the previous terminology we used to call these language tables. So what was decided was a central language table which is now called with a more fancy name, a label generation rule set or an LGR.

And that language table, or LGR, needs to be developed for the root zone all the characters of all the scripts which should be allowed in a label. And the process which was finalized to develop this central root zone LGR was that we start from what is called the maximum starting repertoire. This is the starting point.

So this repertoire already contains a subset of Unicode, for example, it excludes digits which of course are not allowed at top level, it excludes hyphen and equal characters in other languages and oddly contains what are probably letters or can form letters in the language.

And then what needs to be done and what we are doing is we have to have community-based panels which we are calling generation panels; one for each script. And each of these panels will work on their script and recommend to ICANN what are the allowable characters which can form a label at the top level.

So it's the script community which best knows what the right solution is. And we will (unintelligible) the community to guide us going forward. But each of these generation panels is based on a script. Eventually the language table which we have at the root zone is a single table which contains a union of all scripts.

So then what we have is an integration panel, which is hosted by ICANN itself. It receives the individual proposals from each generation panel for each script and integrates that into the larger root zone LGR. So that's the overall process which we are executing based on - based on what the community decided on how to handle this problem. Next slide please.

And this is currently where we are. So we started early this year on this. Arabic script generation panel so Arabic community was the first one to organize itself at a big script community. And they started work on this early this year.

It normally takes a year or a little over a year for a script community to start and eventually come to a level where it can actually submit a proposal for LGR so they're still working and they aim to submit their proposal towards the end of this year or early next year.

In the meantime the first maximum starting repertory was released for all the generation panels to use as a starting reference. It contains 22 scripts; six scripts were missed out so we are going to add those scripts in MSR as well and they will be added by the end of this year with the release of MSR 2.

Chinese generation panel has recently been formed so both Chinese and Arabic generation panels are working towards their proposals which will probably be submitted towards the end of this year or early next year, and that integrated into the LGR and we are expecting the first version of LGR to come out middle of next year.

Next slide please. But that's a very small part of the big puzzle. We obviously need to work on many more scripts. And currently we are engaging with communities for all these scripts to encourage them to form the volunteer groups, the generation panels for the scripts. And once these generation panels get going they will obviously work towards making the recommendation for the LGR.

We've had some initial inroads to many of these communities and we are now slowly activating them and organizing help - to help organize them. Next slide please.

As far as the communications and outreach is concerned - next slide please - basically their communication strategy is twofold. We would like the community to know what IDN program is doing that we don't want to stop there; we also want to go beyond that and engage the community to get involved in the generation panel work and volunteer work which is needed to take this program forward. So we are actually working on both ends.

We've actually been obviously engaged with all the communities at ICANN meetings but beyond ICANN meetings as well we are going into different regions to engage with different script communities to tell them what this program is about and how they can actually contribute into the process.

Next slide please. And also obviously we have a significant amount of many of these materials which are available for the community to learn what the program is about and how they can contribute to it. It is done through the idea of mechanisms from blogs to email lists. We are also now very actively looking at translating many of these materials to local languages to get this information to more people who are relevant as far as generation panels are concerned.

Next slide please. Next slide. So as far as the next steps are concerned we are, as far as the IDN TLD program is concerned we are going to continue to support and engage with community to start more generation panels, continue the work on more scripts.

We intend to release MSR2 towards the end of this year which will contain all the scripts which are currently needed. LGR 1 is expected in June but that obviously it has to be - ICANN is more in a reactive mode, more where LGR 1 can only be developed if LGR proposals from different script communities are submitted to ICANN.

So we are hoping at least two will be submitted by then, Chinese and Arabic. And those will hopefully be included within LGR 1. But again it depends on how - when these communities come back to us.

We're also looking forward to developing an LGR tool, the label generation rules that or the language table, simply put, for the root zone has a very technically complex format because we want this table to be (machine) parsible so that it can be used by everybody conveniently.

It has very formal specification and therefore what we are doing is developing a tool for people to use this specification easily (unintelligible) to it or to - if there is an LGR available to use it to data mine whether the label is valid or not, what are the variants for this particular label for a particular script and so on. And we intend to develop this as open source so that community can use it beyond just the root zone for other levels as well.

We are also - we will be coming back to GNSO. There was a request by - there was a discussion between GNSO and the Board variant working group in London on revising the IDN implementation guidelines. And we will actually be coming back to the GNSO soon with the request to let us know which particular topics need to be addressed in that case, and that's something which we will follow up soon after the meeting in LA.

We continue to support the IDN ccTLD fast track process and obviously continue our communication and outreach to the community. Next slide please.

And if you want to know more details we have two sessions on IDN on Wednesday morning so please come and join us. Thank you very much.

Volker Greimann: Okay thank you. Do we have any questions? I see Yoav.

Yoav Keren: I have several questions. So this was a good presentation that I think probably for many of the people in the room was something that they, you know, it was technical in some level and I'm not sure the understanding the importance of what you're doing.

You know, IDNs is one of the most, if not the most part of the new gTLDs program. And this, as much as I understand it, and I was a part of this work at some point, this group is taking care of a problem of the variants.

The issue about this is that - that I am trying to understand and that's what bothering me because me and some other people from the community have raised the point that this should have been done before the program has actually started - the new gTLD program. And it's still going on; it looks like it's going to go on for a long time.

Now what I want to understand is we have IDN TLDs already been submitted, applications as part of the new gTLD program, already there. Some of them

have already been approved and implemented. If I understand right there is no variant tables for them. So I'm not sure if the people in the crowd understand what this means.

In the long run this is one of the most problematic security problems on the Internet. IDNs are going to change the way billions of people are going to use the Internet. And this is a serious issue. This has to be done much faster. I think it could be done faster. I don't know how, you know, I'm not responsible for the program but I think at least for the major languages this should be done quickly. That's one thing.

The second thing - and you've said something and I want to get a clear understanding about that and that's my question actually. Because for some reason - and I've raised this point several times before, this group is limited to the top level.

Now in some sense the top level goes for a very, very strict process of being approved by ICANN and you cannot enter anything into the root without really, you know, being approved.

That's not happening at the second level. This is even more important for the second level because the security problem is going to come with variants in the second level - and just for people that don't understand what it means is that you will be able to register a domain name, for example, in the Chinese script with a variant that will look - physically will look when you look at it will look exactly the same but in the back it's a different string.

So the people that are going to type it are going to go to a very different Website than the one they thought they're going to go to. This is a big security problem and we've raised it before. So I don't know how can we do that. But this work should be done quickly and also implemented and being a requirement for the implementation of the second level in all new IDN new gTLDs.

So it's very important for me to hear what you have to say about that. Thank you.

Sarmad Hussein: As far as the pace of the project is concerned it has gone for some time. But now we are - we have a very concrete way forward. As I shared, the previous phases, languages and scripts are complex - complex systems and so there was - clarity had to be brought in on how to address that. But I think that process is now in place.

Now - so as far as accelerating the process is concerned we really need to go out to the communities and get them, you know, get them to motivated to come and contribute to this program. And we are now devising a very active communication strategy for that purpose. You may have actually seen some (unintelligible) circulated to GNSO community for that purpose as well.

And we are trying to reach out to the community in asking them and requesting them to get involved for their scripts and for their languages and contribute to this because as the process dictates the community really needs to take the lead on what we can and what we cannot do as far as LGR is concerned.

So through this platform I would, again, request the members of the community to please step forward and volunteer and contribute to this process so that we can accelerate it further.

As far as the second level implications are concerned, though that is beyond what ICANN's scope is but there is - there are these IDN implementation guidelines, which are used to have some in a way recommendations for registries to look at as far as IDNs are concerned and precisely what you were saying any associated security and stability risks are concerned.

So we, as I said, we are initiating that process. We will be coming back to GNSO asking specific recommendations on what needs to be done as far as IDN - a revision of IDN implementation guidelines are concerned and based

on that we will then start looking at the possible revision of the IDN implementation guidelines. So we will be coming back to you and IDN implementation guidelines specifically do address some of the issues which you have raised.

Volker Greimann: Thank you. Edmon.

Edmon Chung: Thank you. Edmon Chung here. I guess you'd expect me to come to the mic on this topic. So can you go back one slide? There's - I guess the June - I guess Yoav would be very happy if LGR 1 is really June 2014, I guess that's a typo, it should be 2015, right.

So - and also on the IDN implementation guideline, thank you - that was part of my question as well. So do you have a timeline for it? And then I have four items that I wanted to mention about.

First of all I've mentioned this for two years now, there is the Project 7 which is the process for which these IDN TLD - variant TLD could be entered into the root and also it has interactions with the AGB.

So this community needs to, you know, as soon as possible you bring that up. I think, you know, this is very relevant for this - the GNSO community because we may have some work to do there.

Second item is the - actually - sorry the second - that was the second item, right? Was I on the first? So the other item that I want to bring up is, again, I've mentioned this a number of times in the past as well and I brought this up to the Council a couple of times.

Last time - last two times the issue of the LGR panels not having an expertise from this community, especially from a policy side, I understand that the work is being done there but, you know, just to alert that hasn't been completed yet, it's still ongoing.

And that relates to a little bit of what Yoav just mentioned in the second level. And this is something new that I want to bring to everyone's attention, this came up in the road show event in Tokyo that was - I was there.

And I think it's very relevant because, Sarmad, you just mentioned the second level registrations is kind of out of the scope of ICANN, that's certainly not the view of the GDD, the Global Domain Division. And there is a significant, you know, this is news to me as well. There's a significant overlap there.

The issue is that from the GDD's point of view it appears that any particular change in the LGR, let's say Chinese or Arabic makes a slight change in the future, which is a possibility, right? That would require every single registry to go through an RCEP process. So there's 100 Chinese registries; each single one of them will have to put in an RCEP just to make a one single code point change. And that I think is not correct.

Obviously there needs to be some communication between this program and the GDD. And that goes back to my question - my recommendation again to say that somebody in those panels, at least as an advisory role, should come from policy aspect because there are kind of the things that, you know, when you work on it you probably didn't realize but they have policy implications. So thanks.

Jonathan Robinson: So thanks, Edmon. I just - if I could encourage you, I mean, you know, we're meeting with the GDD tomorrow as part of the Sunday morning sessions. And it might be worth just raising that and bringing up in that forum as well. Although I'm sure staff will make their own links as well. But it comes particularly strongly from you since that - you're the one making the point.

Edmon Chung: Because it's very alarming because every change in the LGR would basically affect every single IDN TLD registry that has to go in for an RCEP which doesn't seem to make sense.

Volker Greimann: Would you like to comment on that or shall we move on to the next - in the queue?

Sarmad Hussein: If you have time I can make comments.

Volker Greimann: We are - I think we have no one else in the queue so if you can make brief comments we can close the session afterwards. That will be fine.

Sarmad Hussein: So there were four points raised by Edmon. The first one was IDN implementation guidelines and he was particularly referring to timeline around that. And as far as a timeline is concerned we should probably get back to GNSO within four weeks for potential topics which need to be addressed.

And based on how GNSO responds we will then see whether there is need to go forward or not on them. Well, if there is work to be done we will go forward but that depends on what GNSO feels about it. We will obviously request ccNSO for the same thing and request for topics from them as well. And once we have topics we will compile them and start our work.

As far as implementation of the LGR P7 project is concerned this was also discussed in London meeting. And basically our current focus is to really define the LGR and as soon as the LGR is defined then we go into implementation of LGR. And I think there was a - so that's our current focus remains on getting communities mobilized to develop the LGR itself.

As far as the LGR panel policy input is concerned we have been seeking feedback from the individual integration panel member - integration panel and generation panels. And when there is a requirement given from generation panels or integration panel for policy input we will certainly engage policy advisor.

The final point was on IDN tables and review of IDN tables. And so basically as far as the frequency of RCEP process is concerned there are - this was raised by GNSO members in the last meeting in London as well. And we are

taking concrete steps towards addressing them by making some reference point list available for different languages so that people who are using - submitting language tables have some references so that they know that within those references if their tables are within the tables which are published as reference there will be no reason for any changes from ICANN's perspective.

If they go out of that repertoire then we are also trying to define clear guidelines on how they can document that and along with their application so that further rework is not needed. So there's some more work which is going on in that context. And I'll be happy to connect people who are interested in more details with Russ Weinstein who is the relevant staff who's dealing with this.

Volker Greimann: Thank you. Seeing that there's no further questions I would like to thank you for your presentation and willingness to answer the questions that have been raised. And I would like to yield the floor back to Jonathan.