

---

DUBLIN – IDN Program Update  
Wednesday, October 21, 2015 – 10:30 to 11:45 IST  
ICANN54 | Dublin, Ireland

UNIDENTIFIED MALE: Okay. So thank you all. So we'll start the IDN Program Update. Okay. So thank you very much for joining the IDN Program Update Session. We have a [inaudible] full agenda today. We'll start with a quick overview of the IDN program and the progress so far, and then we'll have Marc Blanchet come and share the progress on the development of IDN LGR Ruleset.

We are then also working on developing reference second level LGRs and I will give an update on the status of that particular project. Following that, we have some community updates. We will hear from three new panels which are being formed. These include the Latin generation [inaudible] generation panel, I generation panel, and then we'll get an update on the coordination of duties between the Chinese, Japanese, and Korean generation panels.

So let's go into the first part of the presentation, which is the IDN program overview and progress. So as far as the IDN program is concerned, we are working on multiple projects at this time.

---

*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.*

---

They include projects which are focused on at the top level and a couple of projects at the second level.

At the top level, the large umbrella project is the IDN TLD program, which consists of developing Label Generation Rulesets for multiple scripts for the root zone and an LGR toolset, which will allow us to use those LGRs.

And then based on these, we are intending to embark on determining how IDN variants, which are defined through these LGRs [are able] to be implemented at the top level.

In addition to the IDN TLD program, we also support implementation of IDN ccTLD fast track process, and I will give you an update on that, as well.

At the second level, there are two projects currently underway. We just started the review of the IDN implementation guidelines and development of reference second level Label Generation Rules. We'll give an overview of those and then, finally, conclude with some of the activities we've been undertaking to outreach to the community so that to inform the community about the work which is being undertaken.

As far as the root zone Label Generation Rules are concerned, ICANN and Integration Panel has been working on developing various guidelines and documents for supporting the generation

---

panels. And since the last time we met at ICANN 53, two more documents have been added to this list, which include the requirements for the LGR proposals and LGR proposal document for generation panels to look at as they finalize their proposals for some [commission] to end final valuation by the integration panel.

And MSR-2 has been out since the last update provided and there's been no change on it since then.

As far as the status of different generation panels are concerned, more and more scripts are now becoming – script communities are now becoming active. Arabic and Armenian scripts, these have recently finished their proposals, and they've recently gone through public comment process, and their public comment process has recently [inaudible] now be incorporating the public comments, and based on [inaudible] submitting their final versions, which will be then evaluated by the integration panel in the context of working towards the first release of LGR.

Chinese, Japanese, Korean communities have been working together actively and making good progress. Recently, Latin community has formed themselves. As far as [seated] generation panels are concerned in addition to those which are already seated [inaudible] generation panels have recently been seated. They've submitted their proposals, which have been accepted

---

and announced, and so there's now active word going on this possible different communities.

We've also been working on LGR specification and toolset to have that specification implemented. As far as the specification is concerned, there is a [Lager] Working Group, which is active at IETF, which is working on the specification and considering converting into to a standard [inaudible] RFC.

In parallel, we are developing tools to implement that. We are looking at three different use cases. First is to create LGR use case, which will allow generation panels to go and use this online tool to feed in their data, and that will automatically convert that input into an XML output, which can then be used as an LGR, formal LGR. This is already available in its beta form, and we'll have a more detailed presentation on it later today.

The LGR use case is going to be available very soon, as well. This will allow community to use the input label, as is shown in the diagram on the right there, and the LGR toolset will actually use reference LGR, which is selected; and based on that LGR, provide different variants of that label and also dispositions against each of those labels, whether they're allocatable [or not].

And eventually manage LGRs will allow – the [DAS] functionality will allow to add or subtract to LGRs. So basically to manage different LGR versions internally and all... This tool is going to be

---

made available open source for the community beyond internal use by ICANN, and so as soon as the tool is completed, we will eventually, towards the end of this financial year, also, aim to make this tool available to the general community.

There have been no new evaluations completed since last update, so we still have the same status for IDN ccTLDs. And the 47 strings, which have been successfully evaluated from 37 different countries and territories.

There was a discussion. The IDN ccTLD fast track goes through an annual review. It was a recently opened this review and there was some responses received by the community based on that. There was some discussion on string similarity. The second string similarity review process and its effectiveness, and based on that feedback from the public and the community, the Board actually has asked ccNSO to look into it and see whether that process needs to be devised.

CcNSO has recently formed the [EPSRP] Working Group, which will be meeting tomorrow for this meeting, from what I understand. So that work has started.

As well as second level reference LGRs are concerned, this request, and it was initiated by GNSO to develop some reference tables, which can be used to facilitate or ease the predelegation testing and registry service evaluation process. This work has

---

now been started. Based on the process, which was released for public comments and finalized, we are now at a stage where the initial guidelines are being developed. They will be presented by [inaudible] today and very soon be released for public comments as soon as the guidelines are cleared after public comments.

The tables, the work will start, we'll develop the tables, and these tables will then be released for public comments. Again, each of them with the expert linguistic and security, stability review attached with each language table. These will be done in two batches, which are listed here. And after public comments on each of the languages, language tables, they will be finalized and then published by ICANN. More details will be shared in [inaudible] today.

IDN implementation guidelines, as I already discussed, have been – this work has been started again. The last review of the guidelines was in 2011. There was a call for experts from different organizations, and those nominations have been received, and we had our first meeting to date, so this work has also started, and we'll keep you updated as the work progresses.

So just to conclude, we also actively are reaching out to the community. We've recently redone the whole IDN webpages at ICANN.org, so please come and visit us and give us more

---

feedback on whether you find [inaudible] useful, and how we could further improve these pages to make this information more easily accessible for the community so we now actually have a very short direct link to [inaudible] you can go to ICANN.org/IDN and access all of the information, which is being shared with [inaudible] as well.

And then we've also been reaching out to the community through ICANN meetings, direct updates to SOs and ACs, and presentations that are different for around the globe. And also, through blogs, which we've been working on, and through wiki pages, which we maintain.

If you would like to participate, please write to us. The best way to get to all of us is just an e-mail to [idnprogram@icann.org](mailto:idnprogram@icann.org). So I'll stop here. These are some useful links to follow up. Again, this presentation is publicly available so you can get to them through that, as well.

So let me stop here and invite Marc Blanchet, who will give an overview on IDN LGR toolset.

MARC BLANCHET:

[inaudible] Chair of SSAC in June 2013, we issued a recommendation to ICANN to synchronize the work done by the IDN program. We [inaudible] the IDN-related issues that were

---

dealt with in the Trademark Clearinghouse. Because we found that at that point in time, there were not enough from our perspective or there was not enough synchronization between the two because the matching rules that they were developing were out of synch with sort of [inaudible] well, the...

They call it matching laws, we in Israel call it normalization. Okay? So the normalization [inaudible] they used were not in sync with the organization that was discussed in the IDN program. So we said it's really important for the new gTLD process that these two are in sync and that there is communication.

[inaudible] communication with the Trademark Clearinghouse and tried to get a clarification on the process ICANN is using to keep these two in sync. We got a written response this week, so it took two years and four months to at least get a message from them, a message that we in SSAC have not evaluated yet.

So my question to you is, have you been communicating or synchronized with Trademark Clearinghouse and their rules? I don't think Trademark Clearinghouse [inaudible] in your activities.



---

UNIDENTIFIED MALE: Right. So this particular work is internally owned by a different individual, but I [inaudible] so we can certainly go back and...

MARC BLANCHET: No. Now it's too late. Trademark Clearinghouse should have been resolved and incorporated with an IDN before the whole gTLD process [inaudible], which it was not. We think, but we don't really know because we haven't got an answer. If only you or Trademark Clearinghouse would have said, "Yes, we have been in sync and it will be good." But we haven't heard. [inaudible] and yeah.

UNIDENTIFIED MALE: Sorry, [inaudible] as I understand, the part of letter, maybe [inaudible] name of trademark [inaudible] some letters may be excluded from [inaudible] existing [inaudible] of trademark.

UNIDENTIFIED MALE: No. What is classified as what we call variant in this room is different than what we're taught in variants or equivalents, according to Trademark Clearinghouse. So there are different equivalent rules for [inaudible].

---

UNIDENTIFIED MALE: So there was some details communicated to them and there were responses received from them, as well. This was discussed, I think, a few [weeks] back. And based on that, there was actually a response, which was developed and actually shared with SSAC. And I can certainly follow up in [inaudible] as well.

UNIDENTIFIED MALE: Yeah, we don't have to dial it into here, but I just want to – and I have to – raise [inaudible] synchronization between [inaudible] program. Thank you.

UNIDENTIFIED MALE: Yeah. Thank you. Okay. So, over to [inaudible] part. You want to try this slide?

MARK BLANCHET: Okay. Good morning. Okay. Thank you. Presenting the IGR toolset. So the idea here is with to create a tool to help GR designers to create their LGR, given that the XML format can be complicated and cumbersome for non-XML savvy people. So it's really a Web front end that creates XML files. That's really what you at least at the first level of description. It is really a web front end that creates XML files.

---

It's going to be open source, you can define and manage variants, you could do some validation, various stuff. The project is in three phases, first is the addition tool. So this has been released in August from our side. It's been available for people to use.

The current phase is the second phase, which you're able to validate label and generate variants. This is almost done. We're finishing the little things around the display and the third phase is more for management of LGR effort. For example, merge LGRs do [inaudible] LGRs and stuff.

So I'll go through a quick walkthrough of a how to define LGR. Show you the screenshots and how does it work, and I'm using French LGR as an example since my native language is French, so it's easy for me. So this is THE welcome screen. Okay. And you could import an LGR, an XML file that you already have, and we also support a few of the RFCs that define formats of language. LGRs, if you want.

You could create a new starting from scratch, and we provided one built-in LGR as an example, and as this bit will be deployed, then we can add additional LGRs, built-in LGRS to help people know how they could exercise the software.

And there's also down in your screen, interface language, so this is just the language of the interface of the Web front end. It's not

---

related to the LGR or anything. It's just a front end, and we're currently working with three languages and as soon as it's open source or before – if people want to contribute additional strings for the interface, that's fine for any language.

So when you do create the new LGR, you click on new here, and then it goes through this screen, which essentially asks you for the name of the LGR, and then validating repertoire, right? Which is the base of the repertoire the LGR is based on. So what happens here is you could do there is a menu of the different predefined repertoire there.

While you're editing, you essentially... It's almost ignore this in the sense that you can edit anything in your LGR. It's more when you put –you press the button to validate your LGR that it will look at what you selected as validating repertoire and tells you if your code points are outside, inside, and do the validation itself. So this is really for end of the process processing.

So if you click create here, you go down, and then you have essentially four tabs, which represent essentially the four different sections of an LGR, which is the code points and their variants, references, metadata, and the roles.

For example, here are the references tab, where you can define new references, change, save, those kind of basic work for references. Essentially, a reference is a URL, the name.

---

Here, we show an example in the previous slide of a reference that we just created, and it's... I can go to edit the code point, so I click on the code point edit tab, and then add the code point. Here you could [inaudible] the code point directly from your keyboard with the native code point or with the Unicode notation. You can also define a sequence, which is essentially a combination of [those] separated by blank space. Here an example we just create a Latin small letter “a” as a new code point added to the French LGR.

Then you could edit the code point – for example, defining variants and the management of that variant, which is a type or disposition, comments, the roles, and stuff like that.

References for this specific code point. Here's an example where we add an existing reference to that code point that was previously shown in the previous panel of references, and then you just assign it.

You could define or specify a range of code points, same input, the [inaudible] for the Unicode notation. And then when you define a range, you could actually do exception, which is manually choose the code points you want to include into the LGR, so you could start with a larger range and then move some of the code points you don't want to be in your LGR.

---

The resulting range is shown [below]. You could do a Unicode sequence, for example, here. You could do a sequence in either input format. Here's a shown with Unicode notation that creates the sequence of letter O and letter E, which in French, it actually has a variant, which is a ligature OE, which will be shown just after. So we're actually going... We're now editing by code point and that code point as a variant that we just typed, and then you could add the various information about that variant. Here you show... You see that this is being defined and that the right end type and the roles and things like that.

You could define asymmetry, which is the inverse mapping. Same here. And what I just said before, for validation, which is and you click summary in the front panel, you will see all the details, the statistics, and the various information about your LGR. Obviously, as we go, we are adding more and more information about the LGR.

The only caveat here is it may be more... It more may be compute-intensive, so just don't do summary every second – or ever. Or it could take a few seconds to get the outputs, so just wait.

Current status of the projects. It's available right now on our own servers until we actually provide to ICANN the code, so it's being installed on ICANN servers. So at some point in time, you

---

will get announcement from ICANN that it's available through their servers.

So right now, if you need access to ours, feel free to send me an e-mail. We update the code as it's going stable and mostly in a few weeks, we'll update with the phase two release, which will, you will be able to put a label, and then it will give you all the variants of that label and all the information about essentially validating a level.

UNIDENTIFIED MALE: So we'll move right ahead. If it is not too much trouble, could I [inaudible] to come here because I think this is probably the only [inaudible].

UNIDENTIFIED FEMALE: There are other microphones throughout the room, but this is the best one.

UNIDENTIFIED MALE: Okay. So then let's move on to the presentation on reference second level LGRs. Over to Asmus.

ASMUS FREYTAG: Yeah. Let me do this from there because [inaudible] trying this way for the microphone and that way for the screen.

---

UNIDENTIFIED MALE: I just have a quick question, Marc. [inaudible] based on?

MARC BLANCHET: Actually, it doesn't matter since the too itself is actually an XML editor in some ways with the web front end. So it does have, obviously, some knowledge of Unicode because it [inaudible] of Unicode but the front end itself. It's more of the back end, which has the information about the Unicode versions. And you could specify different Unicode versions when you start. But it's roughly independent of the Unicode.

UNIDENTIFIED MALE: Is it traceable? Sorry. Maybe I was [inaudible] by the [wheel]. Is it traceable what version of Unicode one is using when [inaudible]?

MARC BLANCHET: Visible in which way?

UNIDENTIFIED MALE: In the user interface.

MARC BLANCHET: [inaudible]. Well....



---

UNIDENTIFIED MALE: [inaudible].

UNIDENTIFIED MALE: The XML has a field for the Unicode version, so how is that field?

UNIDENTIFIED MALE: Right. So that field is just a text string, right? So it depends on what you're asking. But what is not visible is the latest version of the Unicode version. The latest version of the Unicode, which is supported internally in the machinery. This is not visible right now.

UNIDENTIFIED MALE: So just in the interest of time, I think we should move forward and take questions at the end, based on the time left. Over to Asmus Freytag.

ASMUS FREYTAG: So I'm going to give a quick overview of the [inaudible] this. Okay. Now it's gone too far. [inaudible] technology. So I'm going to give a quick overview of the work on the second level reference LGR, which are language-based. So this is different from the root work of where the LGRs are script-based.

---

The presentation was originally targeted despite the title of covering early the guideline phase of this, so I will talk mainly about the guidelines for this work. And the guidelines will describe the development process, clarify some terms, and deal with questions of what the target repertoire should be, and how to identify, qualify, and document the relevant sources.

As a starting fund, it is intended to pick up where .se left off who had created the set o 29 language tables and a guideline document. And particularly with the language tables, the intent is to take them as starting point and then do necessary modifications.

The existing starting point tables are expressed in a legacy text format and have no variance and the format is inadequate to express label [delegation] rules. And some of the sources cite for the language content are rather generic like Omniglot and Wikipedia, and the idea is to find maybe some higher quality sources in some cases.

The process of documenting what's required for each repertoire can be challenging for several reasons. Not all languages have a well-established authority with a clear mandate to define things like the repertoire needed for a language. And in addition, even if a very authoritative source exists, like say the [Akabane Francais] whatever for French, what they publish is not

---

necessarily relevant to the task of establishing a repertoire for IDN.

There's very often a notion of alphabet that they will publish, but done in terms of what a native reader of the language needs to know, not what a computer needs to know. So looking at the question of how to define a repertoire, one finds that there are several different concentric circles.

So there's usually an absolutely essential subset that is needed to cover a language, and then there may be, in addition to that, a common use subset that includes a number additional letters because some words in many languages retain some features of spelling of languages they were borrowed from or other purposes.

And it is in particular the common use subset where the authorities are often quite vague as to what is part of it. They don't give you a good answer.

However, there are additional sources one can consult that would provide much better information for the purpose of developing reference tables. One of them is the common locale data repository that is a database that is managed in an open process by the Unicode Consortium, which feeds on input from experts in the various communities.

---

And they have, for all the languages that they support – and they cover all the languages that are of interest for the current project – published a set of essential code points as well as a large set, an auxiliary set. Their particular definition for an auxiliary set is rather maximal, so in most cases, a suitable repertoire for an IDN table might be a little bit less than what is published there.

It is also useful because the task of creating a Label Generation Ruleset is very specific. It is useful to actually look at what has been developed as actual practice in the various languages, and there are some very well-designed IDN tables that should be considered, especially many of the ccTLDs where they publish IDN tables for their own languages or languages native to the territory can provide useful input.

After clarity on what sets to support and the possibility of maybe supporting an essential and common use subset in some way that the user of the reference table can select whether to be more strict or more permissive, the draft Label Generation Ruleset will have to be reviewed and there are several different levels of review that are anticipated.

This slide documents the kind of things that a set of linguistic experts that has been hired for the process will consider, and there will be a separate reviewer that deals with issues of DNS

---

stability and security. This slide lists in detail the kinds of considerations that will be considered for that review.

The deliverables that are intended for this process are for each language consists of an XML file giving the actual Label Generation Rulesets in a machine-readable format. That will be accompanied by a descriptive document that documents both the content in a more human-readable form as some of the inputs that go into it – in particular, the sources consulted. And then attached to that will be the expert report of their review on the LGR draft.

And this bundle of documents is intended to be submitted for a round of public comments where further changes can be suggested and made, and if changes are made, then the idea is to have the linguistic experts re-review the results to make sure that there aren't any new issues introduced. And after that round, there will be a finalization of these documents.

I have in this slide a very rough timeline. We are already behind it somewhat, the guidelines are we're supposed to go to public comment at the beginning of this month. There will be maybe the end of this week that they could go out. Anyway, so by about the end of the second quarter of next year, we should be in the process of finalizing these LGRs. Thank you.

---

UNIDENTIFIED MALE: We'll take a couple of quick questions, and then we'll move on.

UNIDENTIFIED MALE: I have [inaudible] just as an intellectual and academic pursuit. But I'm worried about the wisdom of this being taken up by ICANN. And I see some dangers of micromanagement at second level into which ICANN [inaudible] we should now enter. And I see some dangers going against universal acceptance. This seems to be some kind of assumption here. That second level, you need to specify a language.

You do not. You don't specify a language for ASCII domains. There's no reason why you should specify a language where second level IDNs. And [inaudible] like Arabic script, which is used by many different diverse languages. Now in the new gTLD applications, there was a question by ICANN where you had to specify a language, and we tried very hard to explain to them that it's not the question, there's no specific language that is associated to the particular application we had, and it was no use because they had set up their [inaudible] so that you had to specify a language.

So again, it rings a bell for me. It seems to me that I hope this is not meant to be that we should attach every IDN second label to a specific language. This would be really something dangerous. It will weaken the case for IDNs because then you're seeing all

---

the non-ASCII scripts as something that language is specific, but the ASCII as universal. And this is going against the universal acceptance philosophy.

UNIDENTIFIED MALE: I have no input to that [inaudible].

UNIDENTIFIED MALE: I can quickly respond to that. Basically, at this time, you can specify either a language or a script at the time of application.

UNIDENTIFIED MALE: Maybe they changed it.

SARMAD HUSSAIN: So, again, if you specify a language, then, obviously, these language tables will become applicable. If an applicant refers to a script, then there will be script level tables, which would be applicable. But as has been shared, these are reference tables. It is not required by anybody to conform to these.

I think what is being suggested is that if somebody wants to remain within these reference tables, they can go through the pre-delegation or [inaudible] process fairly smoothly without going into too much evaluation because these have been pre-evaluated in a way.

---

So it's not binding at all and, again, everybody's still open to propose their own tables, then their own subset of this. And we can probably talk about this in further detail, but just, again, in the interest of time, I think we should [inaudible] follow-up questions during the end. So thank you.

We'll move on. So I'll invite Chris Dillon, who's the co-chair of the Latin Generation Panel, to update us on [inaudible] by this panel.

CHRIS DILLON:

Thank you very much, Sarmad. So that serves as summary [inaudible] go through a couple of case studies, actually.

So distribution of the Latin script, it's quite frightening. So the dark green is Latin. In fact, this particular plan isn't completely correct. There are some areas of mixed usage, which aren't indicated. The light green is Latin and another script.

So a couple of case studies I was talking about before, and some very large numbers here. So in Africa alone, 2,000 languages, of which 500 have orthographies. The Latin script has been also extended or modified, so you've got things like tone or accents being indicated, diagraphs, and even quadrigraphs, apparently, in some languages, and actually a pretty similar situation in the Americas, as well.



---

The second case study is just Romanization, and Romanization falls into really two categories. So you've got what I would describe as standard Romanizations, which would include major Romanizations like Pinyin from Mandarin Chinese, but there are also informal Romanizations and here we've got an example of Arabic chat.

Coverage by the panel. So the long and the short of it is that we are lacking representation in many areas, so we are still – we aren't formally formed yet, so we are still reaching out, particularly in Central Asia, East Asia, Southeast Asia, Australia, New Zealand, and the Americas. These are the biggest areas where the coverage is poor.

What are the next steps? So we've only so far just had a couple of meetings. As I was telling you, we're still adding numbers. We're working on a work plan. Lots of study of procedure documents and the maximal starting repertoire.

Then fairly soon, analyzing similar code points, especially the MSR. Also, one important aspect is the coordination with other generation panels and the big one is Cyrillic, but in the future, Greek, some coordination with Armenian, possibly with Georgian. So communication going on there.

---

Into the future, create the repertoire and a whole label evaluation rules. Create the XML and the WLEs and submit for review. So yeah.

SARMAD HUSSAIN: Okay. So that's great. We'll just – if there's a quick question.

UNIDENTIFIED MALE: [inaudible].

CHRIS DILLON: Yes. Cyrillic, there is a huge overlap with Cyrillic. Actually, I was going to catch you later and say if it's okay for me to come to your meeting later for that very reason.

SARMAD HUSSAIN: Thank you very much, Chris. So let's move on. I would like to invite Rapid Sun from Khmer Generation Panel to update us on the work they've been doing in Cambodia.

RAPID SUN: Thank you, Sarmad. So my name is Rapid and I'm from Cambodia, and also my first time fellow for ICANN meeting. And now I would like to report to you about the Khmer Generation

---

Panel. The introduction and the membership and the [inaudible] progress.

My language has been written since early 7<sup>th</sup> Century using a script originating from India. And also, we also wrote some script from Sanskrit and [inaudible] in the current also, we translate any word from English or from French we don't have in Khmer. We also use Bali or Sanskrit to do it.

And also, Khmer was borrowed and found in Thai and Laos and in some ethnic language also, and it is used for 15 million people, mainly in Cambodia, and some [type] in South Vietnam and also in [Thailand].

And the Khmer script is in [Thai] of abugida and also it is derived from Brahmi. And character, 146 character, and the Unicode train from U1780 until U17FA and U19E0 and U19FF.

Here is our member because the Khmer script mainly widely use in Cambodia and other ethnic community, they speak their own language but they also use the Khmer script. So we have about ten community members and they are all in the same county and in the same city, so we are easily to host a meeting or communicate each other.

And currently, about the process of the code point repertoire, we finished all the consonant, [dependent] vowel, independent

---

vowel, and [VeriSign] and [currency]. We already completed analysis.

And we are still working to finalize the version that in Khmer script, there are five form of written, in the middle and two under script, and then the two above the script. And I will welcome [inaudible] some type of character is similar to Thai and Myanmar – only small part, one or two character only.

And here is our next step. We are going to finalize the XML, the XML, [inaudible] Label Rule early in 2016 in [inaudible] by finalize in February 2016. Yeah. Thank you.

SARMAD HUSSAIN:

Thank you. So we'll take the questions, I guess, towards the end. So thank you very much, Rapid. Let's move on to our presentation on Thai Generation Panel. Panus is going to be present here on behalf of Thai [Generation Panel].

PANUS NA NAKORN:

Good morning, everybody. My name is Panus. I'm working at the [inaudible] Transition Development Agency under the Ministry of Information [ICT] in Thailand. Actually, let me state that again because, actually, in Thailand, we have around 68 million people as a total population, but around is 1/3 able to access the Internet, as well. So one barrier might be the languages.

---

That's why as we thought about this is the right time for the Thailand more focused on the LGI, as well. And also, in the Thai government, to start announce on the digital economy as the roadmap for the [Thai country]. This is why we think about this is the good opportunity for [inaudible] system to [inaudible] the Thai people to access the Internet to [inaudible], as well.

And it [inaudible] that the Thai scripts [inaudible] my friends from the Cambodia mentioned about the Thai language is borrow some script on the command so [inaudible] the overlapping or even the variants between Thai script and the Khmer script, as well.

But total right now, we have – I mean, by the 35 languages under the writing system, under the Thai script, as well. If you look at a graph, 20 million people use the Thai language, [based on] the Thai script, as well. But in these proposals, we focus on only the Thai language, [inaudible] is actually first because we thought it might be hard to identify all the languages in the same time, so that why we start in phase one focus on the Thai language for the 20 million people, and then move forward to the northeast Thai and the northern Thai, as well.

And if this is an example that the variants that we try to identify between the Thai script and the other script, as well. For example, in the [inaudible] this might be the variant with the

---

script in Indian, as well, and also it might be for the [inaudible] one, the [inaudible] Thai script might be related with the Laos and Khmer, as well. [inaudible] appearance is quite similar, but in the Unicode it's different. That's why we try to identify what is variance between Thai and other languages.

This is the structure that we just [inaudible] on the September 2015, and we have the Advisory Committee to oversight, make sure that we bring on the stakeholder to join in the panel, as well. We have the Unicode experts. We have the [policy] and [standard] person to join and registry, registrar, and also two representatives from the committee. And finally, we have the language expert, as well.

This is the timeline. We thought this might be finished on the March, maybe March or the February next year, as well. Right now, we already start on the – have some look at the variants between Thai script and the other script, as well. And we're going to hold the public comments, first public comment round in mid-December this year to look at the first draft of the proposal, and we might need to start on the further step, as well.

And this is my [team] and the Advisory Committee. You can e-mail to me, as well. I'm going to keep you [inaudible]. Thank you so much.

---

**SARMAD HUSSAIN:** Thank you very much. And move onwards to last presentation for this morning's session. Could I request [inaudible], who's chair of the Japanese Generation Panel, present on behalf of the Chinese, Japanese, and Korean Integration Panels reporting on their coordination activities, challenges, and solutions.

**UNIDENTIFIED MALE:** Okay. Thank you, Sarmad. Good morning. I'd like to speak about the [inaudible] coordination, challenges, and solutions. I reported this last and [last to] last meeting in ICANN, so this is an update [inaudible]. If there's an overlap with the formal presentations, I will skip them.

This is what I present every time, but I put some example, character example here, so Hiragana, Katakana, and Han characters are used in Japan. And, of course, the Han character is used in China. And Korean use Han and Hangul. Mainly Hangul, but they may use Han. So Japanese GP has to deal with Hiragana and Han. Chinese only Han, and Korean, Hangul and Han.

Han character is shared by three languages, so it should be coordinated how to make an LGR here. Typical issues, it could take character repertoire. MSR has about 20,000, among them, CCP Chinese [inaudible] about 19,000, and Japan [6,000] and

---

Korean 5,000, and many Han characters are shared by CJK. Of course, the big one is the 20, so [inaudible] shared by JK.

And some characters have different uses, meaning in different [languages]. So we have to coordinate and sometimes compromise to make one root LGR for Han characters. And [inaudible] different in different languages. So for example, just an example is they used every time we meet, Chinese defines [inaudible], for example, these three characters. First one is a country and country. And second one a [inaudible] and machine for Japan and for Korea, but in China, the second example [inaudible] both are machine. So it isn't easy to compromise that, but we have to do that.

And [inaudible] things are different from language to language. For example, Chinese characters have mainly two category – one is a traditional characters and one is simplified characters of them. But in Japan, we don't; and in Korea, they don't; and so the root for Chinese characters or [inaudible] string is no mixture of simplified and traditional characters is allowed. But in Korean and Japanese, they're allowed.

So what the rules for such things are? This means that some combination of characters are prohibited in [inaudible] and all combinations of characters are allowed in [inaudible] in Korean.



---

And we have to coordinate. Of course, we know. And last year, we had meetings in ICANN meetings. It was ad-hoc but Coordination Committees, a kind of formal committee, was set up in May. We met in May in Seoul for [inaudible] four times during [inaudible] meeting, and in Dublin, we met twice and we're going to meet tomorrow twice. So more meetings needed.

In conclusion [inaudible] early next year, Sarmad only asks me, [inaudible]. Every time, next time, next time. That's my answer. The framework [inaudible] Chinese, Japanese, Korean. Each GP creates or defines its own LGR independently, and [inaudible] and then we will find the difference or the issues in the combined LGR and we get back to each country and then feedback. I hope this feedback should be in [inaudible] at most three times, but I don't know.

[inaudible] discussion items. Everything is a number [inaudible] variant labels. We always talk with the IP members – [Information] Panel members – about the reduction of [inaudible] characters, and we revise [inaudible] with amended definition [inaudible] rules, and of course, [inaudible] solution found, and we use our real data under .cn (Chinese) and .jp (Japanese). We both have more than ten years' about the second level IDN.

---

So at the second level IDNs and we will find what the problems are when variants of Chinese characters is imported to Japan, what will happen. So using the real data. But we may need some [inaudible].

And the other thing is the investigation of the possibility of using root LGR as a process element to gain more flexibility. We have just begun to talk about this with IP and with ICANN. So we are discussing about how to propose it to IP or ICANN, so propose [inaudible] and send to ICANN, I hope, in [inaudible] or so.

This is the current thinking of the process [in the region]. So at some point, human intervention interaction should solve some flexibility, but I don't know it is good or not, so we have to think about this more. Thank you.

SARMAD HUSSAIN: Great. Thank you very much. This ends the presentations, and I think let's open the floor for more questions and discussion.

UNIDENTIFIED MALE: [inaudible] question [inaudible].

UNIDENTIFIED MALE: Yeah.

---

UNIDENTIFIED MALE: You said that you observed a different string [inaudible] different languages [inaudible] same subset. Do you have some real world example [inaudible] Japan and China of strings that are impossible in one of the [inaudible] versus [inaudible] registers in [inaudible].

UNIDENTIFIED MALE: Yes. We did find some examples. For example, a university name [inaudible] four variant labels, and two of them are prohibited in the original Chinese group. But in Japan, the [inaudible] use this every [inaudible] label. So this is one of the [inaudible].

UNIDENTIFIED MALE: Do you have an example where they use a primary name [inaudible]?

UNIDENTIFIED MALE: [inaudible] that example [inaudible] example. They use every combination [inaudible].

UNIDENTIFIED MALE: [inaudible].

---

UNIDENTIFIED MALE: There's no [inaudible] that until now, whether looked at various [inaudible] and evaluated the possible combinations of the characters [inaudible] there are all possible – I mean, it's always possible to combine all characters. And I see now that we have an example of the Sanskrit used by two different languages that [inaudible] the combination permitted in one and forbidden in another [inaudible]. So probably we... I don't know how to solve this easily because we are already deep in the process, but probably the application process there should be some [inaudible] information. I don't know [inaudible].

UNIDENTIFIED MALE: I think for the root on that is probably a path we should not take. However, there are some scripts for which some languages may use one variant incessantly and some other languages may use some other variant consistently. And in that case, you can write rules into the Label Generation Ruleset that say, "If you pick this variant at the beginning of the word, you have to stick with that same variant every time it occurs. You can't just mix them."

Which is different, interestingly, from the Japanese case, where, for some reason, [inaudible] anyway, that kind of – you have to have one variant consistently throughout is not followed for reasons that are specific to the Japanese language. But there are scripts and there are languages where you can identify such

---

consistency, and if you can, then I know the members of integration panels would very much like for you to make this identification in an attempt to reduce the number of variants that have to end up being allocatable.

If it is possible, that would be an approach. However, I do not want to encourage people to start for the root to insist on making applications language-based. Applications should be based for a specific LGR for the root, but not Russian versus Serbian versus Ukraine.

UNIDENTIFIED MALE: [inaudible] problem I see from all of this is that somehow it is possible to apply for a name that is, how to say, [inaudible], which is not [typical thing] in Chinese, it may be [inaudible] kinds of [inaudible] possibly the Chinese keyboard interface will not allow you to [inaudible] about variance as such [inaudible].

UNIDENTIFIED MALE: [inaudible] begin to think about variance, where you've probably got [inaudible]. We talked about prohibited [inaudible] not so much prohibited, it's about it's that variant is generated of a mix of those characters, then that variant can be blocked rather than allocatable [inaudible] rather than in [inaudible]

---

that it [inaudible] character prohibited language. I just want to make sure that's allocation.

UNIDENTIFIED MALE: We removed the language context here [inaudible] this way [inaudible] when you [inaudible] the language, everything's together [inaudible].

UNIDENTIFIED MALE: I jot down a few questions. Should I go one-by-one or just [inaudible] pretty short ones. I guess just in response to [inaudible] invention, I think, whether it's languages or scripts or anything, I've always advocated that we call it something else, or maybe just call it an IDN tag. I guess the idea there is to bound the [inaudible] repertoire that a particular string can have.

So basically, avoid the mixed script situation. That's why we have a designator identify – we sometimes call it script, we sometimes call it language. In fact, in the new gTLD process or I think in [future] second level registrations, you're asked to provide that identifier so that there are rules that can be created to avoid mixture of scripts. And that's the motivation of it.

And [inaudible] in terms of the second level situation, but I think [inaudible] explained this is for reference only and it's not a hard requirement.

---

UNIDENTIFIED MALE: Can I just say? In the previous application phase, I mean, the gTLDs, that wasn't the case, and that causes a lot of trouble.

UNIDENTIFIED MALE: I don't like it, either. That's why we're here to try and [inaudible]. And then just a quick, I guess, to Asmus. You mentioned batch one and batch two. I probably missed it but what is included and that's...

ASMUS FREYTAG: That was in Sarmad's presentation. He had a slide.

SARMAD HUSSAIN: Should I go back to that slide? So the reason, first of all, they were split into batches was that so that not all the work is released at the same time for the community, and there can be a couple of months between the two batches, and the way the choice was made between the two batches was basically based on the frequency and complexity of the different languages, which are more frequently requested or are, perhaps, more complex work put in batch one versus the batch two.

So that's how [inaudible] organized in that context.

---

UNIDENTIFIED MALE: I'm sorry [inaudible] similarly.

UNIDENTIFIED FEMALE: They are using different scripts. Maybe that's the reason.

UNIDENTIFIED MALE: Sorry. What was the...?

UNIDENTIFIED MALE: Serbia and Montenegro.

UNIDENTIFIED MALE: Serbia and Montenegro. And difference [inaudible] scripts [inaudible] already in [inaudible] if we [inaudible] Serbia and Montenegro [inaudible] using Latin scripts, too.

SARMAD HUSSAIN: Right. So these are basically based on...

UNIDENTIFIED MALE: [inaudible] very close [inaudible].

SARMAD HUSSAIN: So these are basically based on applications. The applications. These are the languages. There's more of the applications we're



---

asking for. So that's a choice. And, again, as I said, the two batches have been divided visibly [inaudible] there's no... Just based on frequency of request and, perhaps, the complexity of the script.

But at the end of the day, all these 29 languages need to be covered, the division was just too physically [inaudible] for public participation in the process.

UNIDENTIFIED MALE:

I have one more, and that's to Chris [inaudible] find one, hopefully. [inaudible] part of the scope to consider Latin [mixing] and [inaudible] from the CJK, we always often mix [inaudible] into the language. And I'm guessing from maybe some other languages may have that, as well. Is that within the scope? And if not, where would that ever be discussed? I remember when we first talked about it in the CJK situation, that was like we got to throw this out until the Latin has even formed or discussed.

Now it's in the formation, is this going to be within the scope? And I'm guessing this is out of scope because that can't be root because it's focused on the root. But I would like to ask that, as well.

---

CHRIS DILLON: We're actually still very much discussing the what's in and what's out, and we haven't actually discussed that particular issue. My instinct, actually, is that if other scripts were to want to have Latin characters mixed like that, then actually, they're almost sort of adopting those letters as part of their script. So my instinct is actually no, but we haven't actually discussed it, so that's only a [inaudible].

UNIDENTIFIED MALE: [inaudible] that you're not actually [inaudible] Latin but [inaudible] negative [inaudible].

UNIDENTIFIED MALE: Really? Aren't A to Z part of Latin?

UNIDENTIFIED MALE: No, they're not.

UNIDENTIFIED MALE: They are.

UNIDENTIFIED MALE: So we are coming – it's actually a different question, and the root, the issue is quite clear. They will not be the mix of Latin with any script, even where that is common on the second level.

UNIDENTIFIED MALE: So on that note, I think we are now out of time. So thank you all for participating. We'll have another session very soon today at 1:00, where we will be talking in more detail on the LGR project itself. So come [inaudible] so please come and [inaudible]. Thank you [inaudible] continue some of these discussions.

**[END OF TRANSCRIPTION]**