
MONTREAL – IDN Program Updates
Wednesday, November 6, 2019 – 15:15 to 16:45 EDT
ICANN66 | Montréal, Canada

PITINAN KOOARMORNPATANA: Okay, let's get started. So, thank you all for joining. This is the IDN program update and we have quite a number of panels to give their updates today. So, without further ado, let's move on. So, for the agenda today we will give the IDN program update overview for 10 minutes and then we'll go on to the updates from integration panels by Asmus and Marc. And then we will go to a series of updates from the GPs, begin with the Latins.

We were supposed to have Greek as well, but because of the logistics situation, he has to leave, so we will miss that. And then we go to Chinese, Japanese, and Korean, and then move on to Neo-Brahmi and then Myanmar.

For the first one, IDN updates, I'm Pitinan, for the record. So, just to recap, for the ones who are new to IDN, the IDN program's objective is to enable deployments of domain names in the local language and scripts for the global communities. It has to be in a secure and stable manner. So, for doing that we have several projects, both at the top level and at the second level.

For the top level, the main one is the root zone LGR, which is the basis of other IDN work. And the second one, we also have the variant

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.

implementations and the toolset for the LGR. For the ccTLD space we also response for IDN ccTLD Fast Track.

And for the second level, there is an IDN implementation guidelines. And last but not least, community outreach and involvement. So, for the root zone LGR process, just to revisit, is the redevelop by the community. Developed by our procedure, the LGR procedure, which is label generation rule set procedures. It is designed to be in two steps, generation panels, which is the script. Users come together to come up with the rules, how to use, how to compose, or use the script to generate the labels properly in the root zone.

Once the panels finish their work, it will be sent to the second layer of the panel, which is integration panels. And after it can be multiple rounds of consultations, if it's already tested for securities and everything, then, it will be integrated into the root zone.

So, for the status of the 28 scripts, that being technically identified, does cover multiple languages in the world. Right now, we have 16 scripts integrated and many of them, as you can see, is reaching the finish line. And we hope by the next time we give you an update many of the green ones will become blue as well.

Moving on. So, after we have the definition of each script, then the variant TLDs, I'll revisit what it is. So, variance in the meaning is something that perceived to be the same as for the users and we can look at it in two aspects. First, the security aspects, which is the example in the blue one. So, if you see the two levels that look epic,

and epic actually looks the same to the human eyes, but behind that they are different.

The left one is come from Latin script and the right one come from the Cyrillic scripts. So, if these two levels exist in the root zone, then it can create security problems to the users. For this, the rules of the E in Cyrillic and the E in Latin and 0435 in Cyrillic has to be defined by both Latin generation panel and Cyrillic generation panels.

Another aspect of variants are usabilities. So, if you look at the orange examples, same script, there are multiple ways to represent the same thing, for example, in Chinese and in Arabic. So, for these levels they look not the same, but are perceived as the same for the users in the same way. All of the yellow highlighted has to be defined by the community that these are variants.

So, the status of this mechanism, how to manage that. ICANN has developed a set of recommendations and it has been adopted by the board earlier this year, during the Kobe meetings. And now it has been passed to ccNSO and GNSO to take consideration and integrate it into their policy development.

And the next project is the LGR tool set. So, we also have provided tools for the GPs and also is usable for registry operator as well to manage the LGR. The main point of this slide is open source and available online. So, if you follow the link on the presentation which could be found on these schedules. You can also use the code for your own use.

Okay. For the IDN ccTLD; since the last update we have two more pass the string evaluation, they're Loas, in Loas script and .eu in Greek. So right now, we have 61 strings from 42 countries and territories.

Moving on to the second level. So, we have the IDN implementation guidelines which is currently, the updated one, is version 4.0, is the revised version of version three, and includes a lot of knowledge gained from the past, almost seven years. This includes the knowledge from the Russo and NCR projects and also how to manage the script from right to left. For status of this project is being asked by the GNSO to give them some times to study the guideline how to implement.

So, right now the status is with the GNSO. And these are the communications and outreach so far. So, we have outreach to all these different regions in the world. We combine UA and IDN when we outreach, and also, we give updates to the community during the ICANN meetings, like this one.

Also, if you want to follow any information about IDNs, you can go to icann.org/idn, and we also support communities and working groups by Wiki pages, and also maintain a mailing list. Okay, so that's about it for the IDN updates. I will pass it to the next one in the agenda, is the update by the integration panels.

MARC BLANCHET:

So, this is an update from last time we did an update, which is Kobe meeting. The reviews we did, the status of what we did for root zone

LGR number three, version number three, MSR, plan for LGR-4, and future work. We haven't reviewed any GP formation proposals, we have reviewed the draft LGRs for our Bangla, Chinese, Greek, and Myanmar. We have reviewed and integrated Hebrew, all the Neo-Brahmi scripts and Sinhala.

We, since Kobe, we actually proceeded with the publication of the root zone LGR-3, which added the Neo-Brahmi scripts, plus Hebrew and Sinhala. That integration required various, that process has various integration and verification desks to create the final LGR, which as you may remember, has two components, the merge and combine LGR and then the separate element LGRs.

The combined one is actually created out of the different element LGR. It was submitted for public comment in April-June timeframe. We got confirmation of review from the various GPs. So, that means that the process we did actually didn't create issues with the submitted LGRs that were submitted by the GPs.

One issue was submitted on the Telugu regarding a rule restricting some Telugu words, so that was actually in the Telugu LGR. But that was kind of noted during our public comment period, the GP agreed and submitted the change to the IP by essentially removing the rule and sent it for public comment. So, we integrated that slight variation of the Telugu LGR all together, and the final publication happened on July 10th of this year. MSR-4 was published on February 2019.

We did the quick review of the new Unicode additions since version 6.3, given that now we got UA to integrate new Unicode versions by the IGF IV. About a dozen code points could be looked into integrated in the new version of MSR, mostly in Arabic and Latin. We haven't done any real activity for MSR-5, however. We didn't receive any requests for adding code points or additional scripts.

So, and over time, over the last years we received some comments from GPs and others about codepoints that may not be relevant for the MSR, so we may entertain removing some of them, which is around 10 in the MSR-5. But right now, we don't see any strong need to look into publishing a new MSR version, so that's currently, essentially, on hold.

Next slide. Duh, that's me. Our plan for LGR-4, we expect the LGR to be reviewed in next months, if the GP actually sends us their final LGR after their public comment period, soon. Could be the Bangla, Chinese, Greek, Japanese, Korean, Latin, Myanmar which would be significant, LGR-4 release. We did defer two LGR that were submitted before, Armenian and Cyrillic. That was because we want to see the impacts on Latin and Greek now that we hopefully get all of them together, then we can actually integrate, finally, Armenian and Cyrillic.

And actually, this week we had some discussion between the four GPs and the integration panel for that purpose, to see the implications of the variance sets between the different scripts. So, that's the plan, hopefully that requires the GP to submit their LGRs final as soon as possible. Obviously, this, in the fiscal year, means that the LGRs for

the LGR-4, the GP LGRs, script for GRs will enter public comment no later than the first quarter of next year. So, that's a call for, hopefully get all these LGR in soon.

Next slide. So, ICANN and IP intend to finalize the active LGRs that were described in the last slide and publish LGR-4 by the end of fiscal year, ICANN fiscal year 2020 which is July, end of June 2020. Therefore, we anticipate that the next fiscal year will see lower level of activities for IP.

We are currently discussing a work model of having the integration panel being inactive for time and then essentially wake up after winter and become active for a fixed date, at a fixed date, to review the submitted LGRs. And this is currently envisioned as kind of a two time during the next fiscal year. And I think that's it for us. Back to you.

PITINAN KOOARMORNPATANA: Thank you, Marc. And then let's move on to the GP updates. I would like to start with Latin, so I would like to invite Mirjana to give us some updates. Thank you.

MIRJANA TASIC:

Now, that's back. Where is the computer? Oh, over there. Thank you. My name is Mirjana Tasic and I'm Latin GP, the last few years. I'll try to be as short as possible. Sorry, I'm pressing the wrong button. Now, next, like this. It will be a very short update. Those are main points which will be addressed during this update. Scope of our work, what

we have done up to now, some updates since ICANN64, and what is our project timeline.

As you can see, Latin script is widely used. And this is a presentation of how many countries are using the Latin script as primary script, or exist with other script. What we have done up to now, we have developed a repertoire, it was a very large task to accomplish, and you can see some fingers. We have started processing languages using Latin script, first we process languages with EGIDS 1-4.

And after that, we decided to add some EGIDS number five languages, which has more than 1 million users. And after one and a half years, or maybe two years, we finally created a repertoire, which has about, let me see, we tested 193 codepoints from MSR-2 and we found a few codepoints which were not in MSR-2 and MSR-3 and those are added. So, completely, we have in the repertoire about 210 characters.

Last year and a half, or maybe two years, almost two years, we are developing variance, it was a very complicated and very demanding task. Hopefully we find a way to work on it and we developed processes and the methodology, how shall we accomplish this. At this moment we are almost at the end of processing variance in Latin script and we have few versions of our proposal sent to IP for the comments. At least it was three up to now, you can see the dates over there.

Now we are expecting to send the next one at the beginning of the February next year. About variance, we find few types of variants. One

of them visual and the other of them are non-visual, that's how we tried to process variance. On the other hand, we have defined cross script variance with Greek, Armenian, and Cyrillic scripts. And also we have other considerations in scope of variance, we have collected basic shapes used in most scripts.

We have considered underlining analysis, and another one, IDN 2003 compatibility problem which we try to overcome somehow. Now, this is our timeline. This discussion which we had before, two or three days with integration partner. I hope we can say that we could finalize our LGR proposal at the end of January next year. We hope it will be done. So, that's all for me.

PITINAN KOOARMORNPATANA: Thank you, Mirjana. And then, next we invite Chinese GP to give us an update. Thank you.

EDMON CHUNG: Edmon Chung, here. Unfortunately, neither Kenny, or Wang Wei, our co-chair is here, so I will try to give a very brief update. So, the Chinese generation panel might have might be the longest standing generation panel, I don't know, I'm not sure. Is it? Might be. But anyway, because we have the biggest repertoire. So the first part is the repertoire. And then of course the variant mapping and then there are a couple of outstanding items which we hope we will complete, hopefully, within the calendar year as well.

So, we started off with looking at the repertoire. From, of course, based on the Chinese domain name consortium experience, plus some of the experimental work that that .asia has been doing. So, the total number we finally ended up with is 19,685 characters in the repertoire. I think we are the biggest but we were not sure, Koreans might be second, at this point.

But in any case, there are a number of different variant mappings that have been updated and revised through the process; I won't go into the details. And this is in the process of the CGP and also the conversations with the integration panel.

One particular thing was looked at, which is to minimize the number of allocatable variant labels, IDN variant labels. Because, Chinese has a simplified traditional Chinese relationship and when you look at it as the name or the domain becomes longer, the variants actually increase exponentially. So, the total number allocatable strings, or allocatable labels become very big.

What we've done, in essence, is to try to minimize that, and we looked at a particular way, I guess, in an engineering way, in looking at the relationships between these simplified and traditional characters as well as the applied for character or the submitted character and made some consolidation, allowing the allocatable variants to be at a minimum.

I won't go into the very details, but what it does is that it narrows it down to roughly, in a majority of cases, about two to three allocatable

variants. So, the way that we do it is actually create a number of dispositions, different dispositions, and use this method to calculate the final, use these dispositions to calculate the final label to see whether they, or to further distinguish whether they are allocatable or not.

We looked at, also from the IP suggestion, we looked into visually similar characters. We followed the Unicode confusable lists and we identified a number of Chinese characters that are potentially visually similar or officially equivalent. We basically worked with a number of linguist experts.

Out of the final six, we've identified that that three of them are distinct enough for local users and three of them are probably very close, but also those are more older characters that are less well used today and therefore we have accepted to map them together. And finally, there is one more aspect that we're working on, which is the out of, I guess, out of repertoire Han characters.

These are characters that look like Chinese characters but were actually created by Japanese and Korean as Kanji and Hanja. And because of the efforts to work together with Chinese, Japanese, and Korean there are certain characters that are taken into consideration for whether they should form variant relationships or not, and also whether they ultimately should be included in the repertoire, in the table, or specifically listed as out of repertoire. That's, I think, the final pieces to be put before we submit, before the proposal we finalize. I think this is the last one. Yes, it is.

PITINAN KOOARMORNPATANA: Thank you. Then next on the agenda, please, I invite Japanese GP to give us some updates and it will be Professor Mori to give our update online as well, thank you.

YONEYA YOSHIRO: Thank you. This is Yoshiro Yoneya from GP, JGP. As Professor Mori explained about our preliminary study of visually similar characters. So, as Edmon said, we also studied the study for the visually similar characters in script of Japanese or variances and we jointly started this work with Waseda University. Professor Mori is a very famous security researcher in Japan and he researched many homographs and such things so, that this, he's a very proper person to do this work.

Okay, so this is a brief background of our research. As you know, the visual similarity has happened in IDNs and it happens also in script, in the same script, and as you know, I'm not sure about Japanese using [inaudible] script. Hiragana, Katakana, and Kanji, and some of them have very similar shapes, so we studied to evaluate the end users recognition of such visual similarity characters.

And in this research, our goal is to develop generic script to eliminate in-script homoglyphs. Not limited to Japanese, but we believe this could be applicable to other languages, but not forced to use. When we studied Japanese in-script homoglyphs confused with characters defined in Unicode. Unicode confuses, is actually confusing to the end users. And finally, we will use this result to JGP's proposal. So from

now, Professor Mori will explain the details of our preliminary evaluation. Can you see to Professor Mori?

PITINAN KOOARMORNPATANA: Professor Mori, you can speak now, thank you.

TATSUYA MORI: Is the mic working?

PITINAN KOOARMORNPATANA: Yes, we can hear you. Please go on.

TATSUYA MORI: Okay, so now we have just started the human perception and elimination of in-script homoglyphs. So, in this slide we are showing 10 pairs we are going to study. These are the pairs of Hiragana and Katakana, and also we have seven pairs from the Katakana and Chinese letter.

Next slide. So we want to ask the questions. The first within the font family affect human perception. We also wanted to study how our word matters. And also sizable letters, and we're also interested in seeing the type of user devices like a PC or smartphones. We also want know whether a linguist background or knowledge affects the human perception, but we haven't just started the first question, does the font family matter. So, we are going to look at the results from this first question.

Next slide, please. So we have gathered the pairs, 10 pairs, using the different font families. In this slide we are showing there are pairs with different font families, but basically, these are so-called San-serif family, Gothic fonts. So what we can see from this is why some characters are almost indistinguishable, red circles, but we also see some characters that have different shapes and also some characters had different sizes. But, these combinations are variable among font families.

Next slide, please. So, this is for another font family, Mincho, which is a Serif font. You also see characters have different properties, depending on the type of font families.

Next slide, please. So, to start with the human perception, a variation. We pick up three, no, no, four pairs of words, as long as they are proper nouns, general nouns, two general nouns. So, here are these pairs of words for human perception study.

Next slide, please. So, we included 16 lab students, both from undergrad students and also graduate students. Mostly they are native Japanese speakers, but some of the students are from China, but basically they are familiar with Chinese characters. And also, this is a limitation of our work, that they have some knowledge about homoglyph attacks, because we as a security research group. So this could be the possible source of bias.

And so, some of the students are really skeptical and tech savvy, so they tend to be really knowledgeable about the differences. So, this is

a task we provided to the participants. So, we provided eight words and the task was pick up the words that they think that they are somewhat unnatural with respect to the composing characters, and the meaning of the words.

Next slide, please. So, we divided it into two tasks, each task uses a different type of family, type of font. And eight words are presented to the participants, while the two, three, and six, and eight are the in-script homilies. So, the task was whether they were successfully identified as possible homoglyphs.

So, these are the results. So the number of participants was 16. So, if we look at task number one and the second word, 13 students out of 16 noticed the difference and but for other words, like 6 or 8, less than half of the students noticed the difference, meaning these words are really confusable for the participants. And we would look at the second task, the status was different.

The number of users, you will notice, the difference increased when we used a different font family, in this case, Mincho. The way you look at a particular word, the eighth word, only two of our participants noticed a difference, meaning this particular character was really confusable even for the [inaudible] students.

So, next slide, please. So these are the key findings we derived from our experience. First, the font types affected the human perception. So, Gothic fonts are more confusable out of the major fonts for our participants. And also, what we note is that in the modern web

browsers like Chrome or Firefox, Gothic fonts are used as a default font, meaning that we are using the more confusable font.

And finally, we found human perception varied among the in-script homoglyphs, so there are some characters that are very confusable, even for the skeptical participants. But, some characters are not that confusable, distinguishable.

Okay, so these are the key findings. And we want to continue this study using a different condition, whether word matters, size of letters, and also, either devices, and a linguist background/knowledge. And we also need to change the set of words. We also had the feedback from of the panel that, in addition to the word-based variation, we may need to work on this study using the character-based evaluation. And also, we may need to increase the number and diversity of the participants, and maybe go to summarize our results and report the results in the future. Okay, that's it.

PITINAN KOOARMORNPATANA: Thank you. Then, let's move on to Korean GP, Professor Kim, please.

KIM KYONGSOK: Thank you. I'm Kim Kyongsok and chair of KGP. Okay. Oh, I see. Thank you. I have five minutes. So, I just will mention important issues. K-LGR covers Korean script. And in this case, Korean script is composed of Hangul and Hanja, Han characters. We submitted the K-LGR version 1.0 in December 2017 and it contains 11k Hangal syllables

and 4758 Hanja characters and 152 variant groups for Hanja characters, plus variant groups composed of Hangal words and Hanja characters, there are five of them.

And we entered into public comment period from January to March in 2018 and summary of public comments made, including Hangal in K-LGP repertoire, it is positive. And then, allowing Hangal-Hanja mixed variable. Several negative comments and some positive comments, and then Hangal-Hanja variant grew, CJK agreement is needed. And some more specific details were mentioned in the public comments.

And next. Most comments were brief, but one person submitted a written document composed of separate pages and we took it into review and discussed. Most of them are accepted in principle and will be deflected in the next version of K-LGR. And regarding Hangal-only labels, Hanja-only labels, and Hangal-Hanja mixed labels, KGP confirmed that there was a general consensus to allow Hangal-only labels and Hanja-only labels, however, KGP has not reached a conclusion whether to allow Hangal-Hanja mixed labels.

Currently, we need to decide two other pending issues, variant groups of Hangal syllables and Hanja characters, and unification for Hanja variant groups between Korea and China. When these two pending issues are resolved, then KGP will divide a calendar year for the 1.0 and will submit to IP. That is it, thank you.

PITINAN KOOARMORNPATANA: Thank you. Then, let's move on to Neo-Brahmi generation panels.

DR. AJAY DATA:

Thank you very much. My turn? Next slide. Neo-Brahmi generation panel is a unique panel actually, where we are dealing with nine scripts, which covers more than one hundred languages in our scope. And our geographical coverage is also beyond our own country where Bangladesh, India, Nepal, Singapore, and Sri Lanka, and another 70 countries use these kinds of scripts. So it's a fairly large exercise which has gone into this panel.

I am the co-chair, Dr. Kulkarni and Professor Singh is also part of us and we are 68 members. So, it's a large panel which we are dealing with for nine scripts. We started in 2015 and we have grown and developed all the LGs. Out of nine, eight have been delegated into root zone on 15 July 2019 and Bangla is being finalized, we are very near to file ideation. 18 months from active work to finish.

We are basically nine panels, kind of stuff, and 18 months of work. We know we have troubled the IP a lot for manuscripts and many updates were sent for feedback. So for the feedback, we were registered in the first round, so we achieved our result very, very quickly and 68 experts as volunteers worked very, very hard on calls and physical meetings, which resulted into this result. That's all, thank you.

PITINAN KOOARMORNPATANA: Thank you. And then, next will be Myanmar.

YIN MAY OO:

So, this is from Myanmar script generation plan now and here's some of our updates. So first of all, I want to share about the languages that Myanmar script covers, which are EGIDS scale above five. So we have Myanmar language, in English it is called Burmese and we have Shan Rakhine, Sgaw, Karen, Mon, and Pa'O Karen. So, all these languages share one Myanmar script. Actually, there are many more languages, but let's concentrate on these.

So, our script is quite similar to Neo-Brahmi, it is based on [inaudible] system, but we have our own we have our own complexity, like the script goes from left to right, basically, but it is not very straightforward like, let it. So, each letter can have their critics up to eight.

So, in this example, is like six or seven critics here. So the bad critics must follow certain order so that the form renders correctly. If the first step and then one them, so that we don't make mistakes when we compute the words, when we process the words.

So, we have medials, which change the property of the consonant and we have several pairs of vowels diacritics and we have other graphemes to change the pronunciation of the first grapheme. And also we have one invisible character that we use to change the syllables into one grapheme. So, that's what's happening in Myanmar script, so the most complex in the complex script. So, our GP has to check all the cross group variance.

So, from our last update from Kobe to Montreal, we have updated more about the letter ‘O’ variants, so we have one console net which can stand alone for us, we pronounce it ‘wah’, so it happens to be a variant of Malayalam Sgaw, Oriya Sgaw, and letter ‘O’ from Latin family. So, the rest are already updated in Kobe. So, for in-script we have many more.

We have single character which looks like the series of three combinations and we have -- so these are just three examples out of many in-script variance, that's why we have to make some rules to restrict what other combinations which can be in the middle, what are the combinations that are supposed to be at the head of the string etc. so that we restrict the occurrence of variants. So, we also have very similar looking diacritics.

So, for Myanmar script, we will use one tag for one script, they can use different type for shell script, they can use another type, and so we take we have taken the suggestion of the IP and to make them into don't-mix rules. So, for example, we can spell wiki in Burmese language. Mon will also want to spell wiki in their own script and they may not look exactly alike, but for people who have only one keyboard may want to access the same thing, so we have looked into these considerations.

And so, we have to classify a bunch of characters so Myanmar script has setting from 1000, we have up to 108F. We have a bunch of classifications to consider: what other ethnic scripts, what are they code-wise, what are their diacritics that is only useful at any scripts, so

that we don't happen to be mixed up with different languages. So in the same script, we can still get the names for what the registrar wants, but we want to make sure that they don't form funny types of graphemes, which do not make sense for any group of ethnics.

So, we have updated up to 15 rules. So, step from rule number 10, they are mostly dedicated for at any languages. That's all we have updated. So, in the future, what we want to update more is to make sure about that, no missing variants and we will be ready to submit our next drops soon, in December, and release for public comment, then we want to submit in early 2020.

PITINAN KOOARMORNPATANA: All right, thank you. So that's the update from the GPs and we have good time for Q and A. So now, the floor is open. Anything online?

EDMON CHUNG: Edmon here. Actually, I have a question that I jotted down earlier regarding the Latin generation panel. What is it meant by compatibility to IDNA 2003, and I guess, why is that needed? Just curious.

PITINAN KOOARMORNPATANA: Dennis will respond, he is the most informed.

DENNIS TAN:

So, if in the Latin script, right, when the protocol changed from IDNA 2003 to 2008, the mapping changed, right? From, we went from a mapping protocol, I mean, sorry for just, you know, oversimplify to a more algorithm fashion, and you have an, in that transition, there were four characters that do not behave the same. And one of it, it belongs to the Latin script with the sharp 'S', [inaudible] the F.

So, under 2003 the sharp 'S' was mapped to a sequence of two letter 'S,' whereas in IDNA 2008 they will leave them as just a different character on its own. But that behavior carries on, on the browser, so you still have browsers, namely Chrome and, I think, IE that still treat the character as in 2003, right? I think what they call this IDNA 2008 + TR46.

And so, because of that behavior, we need to look at that character with different lenses and, you know, whether a [inaudible] solution, you know, is warranted or not, which, in this case, it is. That that's one of the examples.

EDMON CHUNG:

Edmon here, again. I understand the issue now, just a little bit wary of talking about 2003 compatibility. Just a general observation, I guess, if ultimately it's decided that it should be mapped as IDNA variants then the rationale that points back to the beginning of why we explore it might be okay to talk about IDNA 2003, but the rationale, finally, it seems a bit strange, because we're, you know, it's something that we should, at least in my mind, is something that should be obsolete and

phased out. And then, you know, just the, I guess, my feeling about the use of the term. Thank you.

ASMUS FREYTAG:

This is ASMUS, from the IP, for the record. Edmon, I totally agree with your feelings on this issue, as IP member and German speaker. I have two different perspectives on this one, and I can assure you that it actually doesn't matter that this particular character has a history in IDNA 2003, it has other histories in the orthographies of the German speaking populations that would probably lead to the very same solution that the Latin GP has found, and, or to some at least the worry about the solution.

We're not quite done deciding whether, you know, it's the final word or not, but we're still reviewing. But there are differences in the orthographies, which two different countries speaking, unlike Chinese, not just the same language, but reading, writing, everything, except for one character the same way. And that is a little bit of an issue, especially since after the latest reform in Germany, you now have the use of that character linked to the pronunciation of the word which, didn't used to be the case.

The two variants are pronounced by themselves the same, but the preceding vowel now has to be either long or short, depending on which one you use. So it all complicates the issue and there is no easy solution, and I think the Latin GP may have, if my colleagues and IP agree, may have found a good solution for that. Unfortunately, it is

running directly into cross-script issues. That is the real issue in this one.

And there is a little bit of discussion still going on. We had the discussion with the Greek generation panel and the Cyrillic generation panel and, as a consequence of transitivity across scripts, the solution that the Latin GP has found, which is semantic variant, would switch in the transition to cross script to a homoglyph variant. And by having two different types of variants wrapped in the same variance set, you get cross script variants that don't look intuitive, but they may be required to be secure.

And it sounded from the discussion that the Greek and Cyrillic GP were amenable to live with that situation, especially since it only affects a single variant pair. But it is definitely an interesting case, and I think you have good instincts noticing that particular one that's worth discussing.

PITINAN KOOARMORNPATANA: Any more questions from the floor?

MIRJANA TASIC: I just wanted to point one. Yes, we lost two months on this, every week one hour.

PITINAN KOOARMORNPATANA: Okay, please.

MAHFUZUL MAJUMDER: Thank you very much to everybody. It is not a question, just a comment. I'm Mahfuzul Majumder from Bangladesh, for the record. You know, language and script related issues are very much sensitive to the Bangladeshi people, especially the language movement of Bangla that took place in 1952, is the very spirit, first spirit, of Bangladeshi which led the nation for the independence in 1971.

For this, even when we had been working to developing the LGR for TLD, we were very much careful working for and we scrutinize all possible ways to make the LGR documents that representative and protect the basic principle of Bangla language. It is true that we agree with the NBGP and due to a very narrow scope, especially written in TLD using [inaudible] based alphabet. Consequently, our country gives importance and seamless access to the internet for most people.

As a representative of NBGP, Bangladeshi work group, and as well as the GAC member of Bangladesh, I'd like to express two points in this session. I would like to submit our significant concern on LGR for the second level domain that I have already discussed with the co-chair and also Dr. [inaudible].

It is our humble request, please involve Bangladesh before starting the development activities for the second level domain, and additionally, we are letting you know that Bangladesh will continue its activities with the Unicode consult team currently to solve these issues. Thank you very much.

PITINAN KOOARMORNPATANA: Thank you very much. Thank you all. If I am hearing none, question online or on the floor, so I will close the meeting. Thank you all.

[END OF TRANSCRIPTION]