COPENHAGEN – ICANN GDD: IDN Root Zone LGR Workshop
Wednesday, March 15, 2017 – 09:00 to 10:15 CET
ICANN58 | Copenhagen, Denmark

UNKNOWN SPEAKER: Okay, I think we should probably start. Sorry for starting a bit late. I just think an early morning for people who are just waiting for everybody to get here.

So, welcome to the IDN Root Zone LGR Workshop. This workshop is largely aimed to address members of different generation panels, about give a chance for generation panel members and integration panel members to interact with each other.

And today, we have overview of, we're in the process of finalizing the second version of the root zone LGR. We have Asmus Freytag, who is going to, who is a member of integration panel, who is going to give you an overview of what is potentially going to be inside the label generation ruleset, and how it's organized.

Then we will go on to community updates from Chinese generation panel, Japanese generation panel, Korean generation panel, and Thai generation panel, followed by question and answer session. We would want this to be a more

interactive session. We are members of different generation panels and integration panel, give an opportunity to discuss various issues which are coming up. And then without further delay, let me hand it Asmus to start with the first presentation on root zone LGR.

ASMUS FREYTAG: All right. So, add to the introduction. Originally when we conceived these presentations, this was to be in the second half today. So, all of you would have had a chance to go and listen to the announcement of what the status is of LGR 2. So, that is kind of assumed in this presentation.

So, we are in the integration panel, pulling together a number of script proposals for root zone LGRs, and we are in the process of finalizing what we call LGR 2, which is the next update of the root zone label generation rules.

So, which of these many buttons do you want me to press? I'll just ask you for the slides. So, in LGR one, which is about a year and a half ago, we started off with a single script Arabic, which we started off with because it did not have any other scripts. It was similar enough to possibly conflict with, and since then, we have been busy with the generation panels to work on finalizing Georgian, [inaudible], Lao, and we are still finishing Thai and

depending on the timing, we may even be able to pick up the Ethiopic script.

Just to give you a visual reference, I have put some samples of the various scripts on the right. All of them, except for… All of the next scripts listed up there, except for Georgian, are Southeast Asian scripts, and they are what we call complex scripts. Kamer, Lao, and Thai, we need to consider together because there are very simply closely related and have some similar issues, and we want to make sure we handle them the same way.

There is also another finish that script LGR that exists that's Armenian, that was entered into a deferred state at the time of LGR one, because we want to wait for Latin, Cyrillic, and Greek to be able to consider all of those four scripts at the same time.

Georgian is an European script, but the overwhelming consensus is that it is not closely enough related to any of the other European scripts to require it being considered together. So, we will proceed with it, LGR 2. And this is not the talk where we talk about future direction of the LGR past LGR 2, so I will skip that and go to the next slide.

We went by two, I think. Okay. So, a key concept to bear in mind is that while I'm sitting on the integration panel, and the process of creating a LGR is called integration, this does not mean that

the result is a single file somehow. So, in fact, the LGR will be distributed over a number of files.

The first one will be an overview document, which is just a text document describing the background, describing the process, highlighting relevant items, etc. The next document is a set of co-tables, which give a graphical overview of the repertoire, and then an example for, in this case, Georgian, has been depicted on the right. You see things show up in different colors. White generally means code points that are not used for, as part of INDA 2008.

The reddish color is the kind of code point that did not make the cut-off for MSR 2, the maximum starting repertoire. And the green and lavender reflect the choices made by the generation panel. So that in this case, the generation panel designed it to include most of the available points, except four of them, which are shown in lavender here.

Now, corresponding to this graphical overview of code tables is what we call emerged LGR, which has an union of all of the repertoires. It has an union of all variant mappings, however, there are all set to type locked for the merged LGR.

And it has a common set of whole label evaluation rules and actions. There is also, for each script, what we call an element LGR. That has a repertoire corresponding to that script. It has

**EN**

the variant mappings appropriate for only that repertoire. It has, it assigns variant types on a per script basis, and that can be something other than block. It can be allocate, and it contains WL rules and actions that can be triggered by code points in that per script repertoire.

In addition, it contains the default rules and actions for the root zone, as defined in MSR2, whether or not they're applicable to labels in that script. I get the next slide please.

Each element LGR is derived fairly directly from the LGR proposal, that is submitted by generation panel, assuming that it has past, you know, public comment, review and also review by the integration panel, and is found to be acceptable. So, given that it's made to cut, then the IP will do some slight copy editing on descriptions and annotations in the file. And retain the repertoire variant assignments and rules, [inaudible] change.

That will become an XML file, and from that XML file is generated an HTML version containing the same information, but being more readable to the human reader. The description of the LGR, is not incorporated into the integrated into the root zone LGR, but it is referenced. So, all of the LGR proposals remain archived, and all the discussion of vice or things that were added, can be found in those archived documents.

**EN**

A merged LGR is a file that is created by mechanical merge, as described in the previous slide. It has a common overall description of the LGR added to it. The annotations for each code, for example, are replaced by pointers for the element LGRS, so when you look at the merged LGR, you can look up where each code point comes from, and what document defined it.

And all the rules, glasses, tag values, and other elements in the LGR are renamed by using a script prefix, so that we avoid name collisions. So, can I have the next slide please?

And in addition to these files, and I talk a little later about how we actually use the different, you know, the element LGRs and the merged LGR, we need to point out that each generation panel is requested to submit test labels, that can be used to verify its LGR.

This is a really, really crucial piece of the puzzle. We really need to have generation panels providing test labels that both labels that are supposed to pass the LGR and labels that are supposed to be rejected by the LGR. So that, in case there is any copying mistake in creating the element LGR, we can run the set of labels and we can prove the skip the same results with the element LGR that we got [with this?] proposal.

If the test labels are insufficient for that purpose, then we cannot guarantee that there aren't any changes.  So, the IP will perform a mechanical verification every time a LGR is created to make sure that all element LGRs give the same results of the proposal LGRs.  And likewise, we can mechanically verify that the element LGRs were correctly merged into the merge LGR.

And in many cases, we can also use the test labels to double check that the common LGR contains the right code points and rules.  After we've done all of our verification checks, plus a number that I haven't mentioned here, the LGR will be submitted for public comment, which allows community members to do further review and checking.  Also allows the original generation panels to double check their LGR to verify that the integration was successful.

Can I get the next slide please?  So, now we come to how would one use a root zone integrated LGR?  In order to illustrate this, let's take some idealized steps in processing a label application. The actual process may well be somewhat different, but you know, these are kind of logical idealized steps.

Let's remember that each application for a label defines a script context, so you get submitted a label, string, and the script for which the label is intended.  All labels in the root will have to be in a single script.  And for the purpose of the root, we are

ICANN 58
COMMUNITY FORUM
COPENHAGEN
11–16 March 2017

counting the complex writing systems of Japan and Korea each, in a single script.

The next step is to process that application, by using the submitted script value to select an appropriate element LGR, and that process results in determining the validity of the applied for label.  The, by executing the LGR, the check was performed, was that if the label fits the repertoire subset for that script, whether all the defined context rules and code points in the LGR are matched by the label, whether the label matches any other whole label rules, and whether there are any code points that have what we call, reflects a variance that mark them as not being part of the repertoire.

The last past needs, perhaps, a bit of explanation.  If two script LGRs are in need of defining variance that go across script boundaries, for example, the Latin and Cyrillic would need to do that, then the targets for these mappings are, of course, outside each repertoire.  And they need to be specially marked when they show up in the LGRs, so that no one can apply for labels that are consisting entirely out of targets for out of script variance.

So, the ability checking checks that, and let's go to the next step. And now once we know a label is valid, then we switch gears and we use the merged LGR, because that's the only table that is

able to be used to check for collisions between labels. So, normally what you do to check collisions, you look at each variant set defined in the merged LGR, and based on the information from that, you can create what we can an index variant for any label, which is based on the merged LGR.

And the two label set have the same index variant, end up being in collision. So, you do this computation for all, ahead of time for all labels that have already been delegated, and when you have an application, you do that for that label, and you compare the index variance for a match, and if there is a match, then the applied for label, plus all its variance, are in conflict with existing delegated label.

And an application process that would typically lead to the rejection of that validation. And then there is a final step, once you have validated the label and you have checked that there are no collisions, you go back to the element LGR, and you generate a list of all possible allocable [inaudible], in case the script contains allocable variance.

And beyond that, there are all sorts of other steps in an application process that are outside the scope of a LGR. For example, the choice of which of the allocable variant labels will actually be delegated is not answered by the LGR process. Next slide please.

So, that is an overview of the basic things that you can expect out of LGR 2, when it comes about. We had originally hoped to have it ready in time for this meeting, but we found some extra due diligence we needed to do on some of the scripts. And we want to err on the side of correctness and completeness then to rush a result.

So, LGR 2 will happen after this meeting, but here, you have a kind of a preview on it. And now would be the time for anybody to ask any questions regarding this part of the process.

I think it's too early in the day. You have a question?

UNKNOWN SPEAKER: So, I think it would be useful for the community to understand, why a split LGR in multiple files, and not possibly have LGR, all scripts [inaudible] into a single file, and you know, what are the advantages we get out of perhaps dividing it up into these parts?

ASMUS FREYTAG: While the simplest answer to this question is that, somewhere we need to identify what are all of the code points that are possible or available for a label in a given script. So, that's a list of code points by script. And it is just given the formats we have for formally describing LGRs, in particular, RFC 7940, it is

ICANN 58
COMMUNITY FORUM
COPENHAGEN
11–16 March 2017

**EN**

convenient to have a set of files, one per script, that gives you that repertoire.

And we also have, for variant mapping definitions, the fact that even in the case of overlapping repertoire, like the Chinese characters, different LGRs, like Japanese, Korean, and Chinese, may assign different variant types to the very same mapping. And again, it is just more straightforward to have the script specific pieces of information in a script specific file, rather than having a very complicated scheme of having multiple pieces of information all tagged by some script tag.

That is a very messy thing, and it's not supported by RFC 7940. So, it's… If you think of the root zone LGR as a database of information, we have just chosen a particular database design that works well with the kind of tools that we have.

EDMUND CHUNG:      Edmund Chung here. I think that makes a lot of sense, and it is not impossible to devise a LGR with all of the different types of variance to satisfy a single list, but well, a single XML. However, I think it makes sense because both for the root and actually this is relevant, I think, for the second-level registrations as well, and we'd encourage TLD registries to adopt similar approach to have an LGR for each of those scripts or language.

So, based on the tag, or the language script tag, for a particular registration. So, you know, I think that makes a lot of sense, even though technically I should say, it is not an impossible device scheme. It makes it much more complicated.

ASMUS FREYTAG: If I may add to that, it is actually curiously of interest that, of all of the files in the scheme that we have, the one you could possibly do without is the [merged?] LGR, because some of the processing steps that we in fact, used the merged LGR, you could fake the results by… For instance, in the merged LGR, all variant mappings is set to blocked, which you need for collision testing.

Well, you can read a normal LGR and pretend all types are set to block. So, it's clear that you could, if you really wanted to, you know, create a workaround, that would do completely without the merged LGR, but we looked at that, and we felt that is less than an intuitive in some ways, in having the merged LGR as a concrete file, provides a very good check to make sure that we understand where the repertoires are overlapping, and the, what the union of these repertoires is, rather than having that idea be just represented operationally by, you know, calculating an union on demand from the element LGRs each time.

So, we think that we have found a good compromise here, and we have something that allows, for instance, a generation panel that has submitted a LGR, to go and double check that it is correctly represented not only in the element LGR, but also correctly reflected in the common LGR. And if it passes that test, then we can be better assured that we have done the right thing, and we understand what we're doing.

UNKNOWN SPEAKER: Any other questions?

WALTER: Yeah. This is Walter from [Idea?] Registry. I just wonder, are there any timetables that could be predicted for finish or for [inaudible] work, enter the next step you mentioned for, you know, how to set up the detailed process to determine which variable string had been dedicated?

Because all of the users feels very, you know, curious about the procedure. Thank you.

ASMUS FREYTAG: Well, that particular question is above my pay grade, so I'm going to have to pass that to someone.

UNKNOWN SPEAKER: So, as far as when the work is going to be completed, I think it's incremental work. LGR one is already out, LGR2, as you already saw, would have four or five more scripts integrated into it, which is going to come out fairly soon, within the next few weeks. And then we will keep integrating as we receive more script proposals.

So, in a way, it is also up to the community to push forward and complete their proposals for us to integrate. So, that's the first part of the question you had. As far as the implementation of that is concerned, we are already currently looking into possible solutions on how these, how we would implement it.

There is homework which is already done, that's being discussed on what is the right possible solution. And as soon as we can find a way ahead, we will come back to the community to get community input on whether that seems a feasible way ahead, based on community feedback we'll finalize that process.

So, that's already in the works, and should also be coming back, we should be coming back to the community, but that may take slightly longer than a few weeks, but in any case, I think that's maybe a few months, but that's already under process as well. Thank you.

And if you're specifically are talking about Chinese community, I think we move right on. And we have Edmund Chung here, who

is going to give an update on behalf of the Chinese GP on where the generation panel is, as well as their work is concerned. And maybe also, a comment on, you know, some of the timelines, question around timelines, which you had around this work.

So, I will request Edmund Chung to present an update on Chinese generation panel work.

EDUMND CHUNG: Thank you so much. And I'm here on behalf of Wang Wei and [inaudible], who are the co-chairs of the Chinese GP. First of all, I apologize. I probably have to come in and out very soon after the meeting, after the presentation, but I believe Wang Wei is online, and should be able to take questions further at the end, or add to what I want to say, what I'm going to say.

So, next slide please. So, the CGP, the Chinese Generation Panel, has been working for quite some time now, probably for three and a half years or so, looking at the repertoire of Chinese characters or Han characters, one of the things that I want to highlight here is that we started off with the Chinese domain name consortium, core set of the characters, because of the integration with Japanese [foreign language], and we also realized that the Korean [foreign language] has expressed interest to be represented in the root LGR.

**EN**

We also looked at commonly used Han characters, coming from the Hong Kong supplementary character set, as well what is called II Core, which are the commonly used characters in the Japan, Korean, and China shared set of commonly used characters. So with that, we have created a full repertoire that represents the Chinese characters that are being commonly used across the three language groups, but then focused much more on the Chinese usage, of which, just a note here that we did identify two characters that is not currently in the MSR.

The master repertoire. So, the maximum set repertoire. So, we will be going through a process to suggest the IP to add those two. Next slide, please. So, what we have found, as you can see, we started off with the core, and then we looked at the Japanese [foreign language] and the Korean [foreign language]. We've looked at the overlap, especially focusing on the overlap between Japanese and Korean, to focus our efforts to make sure that the variant definitions are consistent across the three languages.

Next slide please. So, what we did, part of the basic findings that we based most of the work on this CDNC variant mappings, we identified a number of about 100, less than 200 characters that are used from, they're non-CDNC characters and we worked at looking at whether there need to be variant mapping considerations.

What is called dot Asia characters is actually dot Asia. A few years ago, we experimented with the addition of the Hong Kong supplementary character set, which is a number of Chinese characters that are commonly used in Hong Kong, but not necessarily in mainland China. And so, there were a number of characters that were used.

And so, in the overall review, that was also taken into consideration, to which is the finally, the variant mappings that we reviewed. Next slide. So, over the last two years, one of the key progress that was made is the… Previously, in the consideration of CJK, [Hung?] characters, Korean had expressed that they were not going to use the [foreign language], that has changed in the last two years due to the language actually development inside Korea.

And so, it became apparent that the overlap of the Han characters between the Chinese characters and the Korean [foreign language], needed to be looked at. Over an [inaudible] process, from throughout last year. In fact, I think it was 2015, or that we kind of started looking at it, but over the course of 2016, we have been able to identify the conflicting ones, and narrow it down to about 100 or so, and looking at each of the characters, basically, whether they still should form a variance set or could reasonably be split up.

**EN**

So, we looked at it with a number of criteria. The operational experience, from actual registrations, looking back at the usage of those characters, the semantic rationale and drawing upon linguist support to, from both the Korean side and the Chinese side. And the Chinese side including from mainland China, Hong Kong, and Taiwan.

And eventually, we're quite excited to report that early this year, the meeting in Beijing, we are able to come to a consensus and come to agreement to have a consistent set of variant definitions. So, next slide please.

And just as a note, this is, the resulting set has a little bit of less than 20,000 characters in it. Those with, what is called a size of variant group one, is actually there is no variance. So, more than 60% of the characters, about of the about 20,000 characters, actually do not have variance.

So, that's an important note, I think. And if you add in the second one, which is just having one other variant, usually the traditional Chinese and simplified Chinese, that adds another 20 odd percent, so we are looking at about 90% of the characters, in fact, fall within either no variance or just one variant.

And of course, there are ones that have a little bit more, and you're looking at the numbers. And the most we see today is there are about eight, actually there is eight variant included.

And that represents really just two groups, two variant groups representing 16 characters.  Next slide.

So, through the process, we have been having interactions with the IP as well, because the Chinese actually define what is called preferred variance as well as other variance, we have different types of variance and subtypes to note which character, when the label is formed, whether they should be allocable or blocked.

Under the advice from IP, we actually added one, or neither, to indicate the non-allocable reflects of character, to make it more clear.  So, in the coming version, we will add that subtype.  Next slide.  Okay.

This is simply a note on the actual technical implementation. And showing the [inaudible] being used on certain characters, and I guess I won't jump too much into detail.  Wang Wei, if you want to bring this back up later, please add to it.  Next slide please.

Here is another important statistic for people to consider.  In terms of creating many allocable variance, there has been a consistent note from the IP to try to avoid.  The Chinese Generation Panel has spent quite a bit of time looking at containing the number of potential allocable variant labels that are being generated from the LGR.

And this version, we have been able to narrow down to, for the simplified, you know, mapping to the simplified Chinese characters. Only four characters now have multiple allocable variance. In the traditional Chinese, that is about 128, well 130 characters that will have more than one allocable variant.

That is down from, I believe, about 180 or so from before. One way, you probably have a better number, but we've looked at those sets and been able to reduce the cases where there is multiple allocable variance. Next.

And just, again, the technical implementation on how the label generation, the LGR works. When we calculate the allocable or blocked disposition, we use the whole label evaluation rule to identify which ones should be allocable and which ones should be blocked. Next, please.

So, the next step is, we are looking at, I guess, a relatively large set of characters, compared to many other languages and scripts. It's 19,700 characters, a little bit shy of 20,000. We have received note from the IP to try to reduce that as much as possible. But I did want to bring up one point. In terms of Chinese, it's a little bit different than alphabetic based languages.

These are actually words. We are actually talking about 19,000 words in Chinese. So, if you look at it, a contemporary English

dictionary, you see about 200,000 words in the dictionary. And we're already just 10% of that. So, if you take a look in that context, this is a very, very small number.

It's a tremendously small number to be considered. So, the other part we are looking at is, of course, limiting the number of allocable labels from the LGR, from the list of IDN variant labels generated. There has been different consideration, and I think, ultimately there are three areas that I want to highlight.

One is, we are currently actively looking at the actual usage of Chinese characters, especially in names, especially in domains names, what the frequency of these variants will actually happen, and what the root should expect. Is it a big issue? Is it…?

As I mentioned, more than 60% have no variance, actually more than 90% have just one variant. And with that, does it calculate well in terms of some exceptions could actually be tolerated at the root. The other part we're looking at, is there may be potentially require some policy intervention in the limitation of, not number of allocable labels, but actually limit the number of actual of delegated IDN variant labels into the root.

That may be an approach that is more realistic, rather than to arbitrarily do something in the LGR process. And then the third area that, Wang Wei in the team is looking at, is [inaudible],

which is also the last slide, is potentially doing a, tweaking the LGR to make most of the advanced blocked, but rather than running through the LGR once, running through it multiple times, so that the original label actually remains being allocable and use this methodology to work around the issues, so that we can contain the number of allocable variant labels to a definitive and small set.

This is just one of the proposals, and I won't add too much to that. And that is the end of the presentation. Thank you.

UNKNOWN SPEAKER:      Thank you. We can take one question. If none, then we can… Mark has a question.

MARK:      Mark [inaudible]. Maybe more a comment, but the, as you said, there is the need for the policy development process for those kinds of additional steps. And you know, the current process doesn't… We cannot take this into account, into the current GR integration and all of that stuff.

So, we need to be bound to the current scope of the work.

EDMUND CHUNG: Edmund Chung here. We're well aware of that, and as I mentioned earlier on, the CGP is working very hard to convince, I guess, the IP that sets we have right now is reasonable, and if there are certain exceptions, perhaps it needs to be tolerated as exceptions when you look at the majority, even maybe super-majority of the cases that generates a reasonably small set of allocable variance.

But we're well aware of the point you made.

UNKNOWN SPEAKER: Thank you. So, let's move forward. I will request Hiro Hotta, the chair of the Japanese GP, to please take us through the update from the Japanese Generation Panel.

HIRO HOTTA: Thank you, [inaudible]. My name is Hiro Hotta from the JGP, Japan GP. And I'll, very briefly, update the Japanese GP status. This is an event called [inaudible], skip please. Yes, okay. So, as [inaudible] said, that the Japanese, Chinese, Korean, these three LGRs we share thousands of characters.

So, we need to coordinate. So, about the coordination, Edmund explained well. So, I would like to skip this chart as well. And this chart means that this is a kind of procedure for three GPs to take, and we are thinking about the Chinese, Japanese, and

Korean LGR independently first, and then if there seems to be conflict between us, about this year, the script, which is [inaudible].

So, we need to coordinate about that, and we merge the variant definition, and then merge or coordination of the variant definition, that's the, I think that's the hardest part for three of us. As Edmund explained, or reported, the coordination has a final result now. So, I'm very happy to be in this procedure goes to the right hand side.

So, after merging that, the Chinese, Japanese, Koran each LGR will take the variant definition from other original, other script LGRs into each own LGR. And then we'll think about that, whether it works for each community. Next please.

Yes, this is a size of the [inaudible]. We have 6,358 characters, maybe one or two will be added to this, but almost 6,4000 Han characters. We at [inaudible], different separate scripts from Han characters, but we Japanese, everything use the mixture of Han [foreign language].

So, these are the three scripts which will be incorporated into the Japanese LGR. And for the variance, we, the Japanese GP decided that Japanese LGR originally have no variance to be defined, but we're input the variant definition from CNK. And

ICANN 58
COMMUNITY FORUM
COPENHAGEN
11–16 March 2017

about [inaudible], rules under discussion, but the size of the rule is very small, it should be very small.

And maybe, the, for the reduction purpose, reduction of allocable labels purpose, we're using temporarily. Next please. So, development at and after Hyderabad. So, we have two [inaudible] JPG. JPG has two big issues, which is under about the variant definition. So, as I said, JPG waits for input the Chinese and Korean JPG definition of variant characters.

So, we wait, we have waited for the result of their coordination, and as Edmund reported, it [inaudible], or it has a result now. So, we are thinking about that, how to incorporate it into our LGR. That's the first thing. So, the first one will be moving rather fast from now on. And the second one is at the bottom of this slide.

IP [inaudible] to reduce the number of allocable labels. So, because we input all of the definition from C and K, and Japanese word has no preference about the combination of the characters, which means that all of the characters are independently use in a string.

So, it means that if a string has a lot of variant, it may have the variant, the number of variant labels may be a big number. For example, 10,000 or so. So, we should try some ideas to reduce

allocable labels.  So, we are thinking about his now, and JGP discuss about this with IP on this Sunday.

And had that kind of hint to [inaudible] that into, to reduce the number of allocable labels into very small number.  So, I will be able to report that in the next ICANN meeting.  Okay, next please.  And this chart is very much like, what Edmund said, in the last, in his last slide.  And we also thought about this method, which is a parallel execution of LGR to reduce the number of allocable labels, but for the time being, we consider that the reduction of the number of the LGR level is important.

So, for the time being, we forget this to be proposal or requested to ICANN, or IP.  Okay, next please.  Yes, that's all.  Thank you.

UNKNOWN SPEAKER:    Thank you.  We have a question.

UNKNOWN SPEAKER:    [Inaudible].  Can you [inaudible] your slide, let me see.  Keep going.  Here.  Okay.  Next one.

From my point of view actually, when we are talking about IDN variance, the real, the variance as it came in from the Chinese characters, because they simplify and traditional one, all right, there is a maturity that I think there is [inaudible] most of the so-

called variant initiative. It becomes a Chinese character, a simplify and traditional one.

I think we might be… I don't know if that would be easy or not. Maybe we can make this a very much easier, is because in Chinese, very few of the name, you know, the TOD name, would be combined with simplify and traditional together. Either it's all traditional, or all simplify, but right now, I think it seems like we try to solving the problem much bigger than that.

We almost, we're also thinking about, what about the people trying to make, mix the traditional and simplify in one stream? So, that is reason why this is trying to solve the problem, very big. It can be much simpler if we're just solving, it's all simplified Chinese character stream, or all traditional Chinese character stream.

So, if in that case, the [inaudible] you [inaudible] here, instead of using Han character in Chinese, community is one big cycle, might be, should be a two cycle. One is a simplified another one is a traditional one. Because that is really what the variance came from for CJK.

I think you agree, right? Because most of the trouble of variant is actually [inaudible] simplify and traditional Chinese character.

UNKNOWN SPEAKER: Yes.  I think I understand the issue, and CGP itself is trying to solve that, to reduce the number of variance, allocable variance, by thinking about the traditional and simplified and new mixture.  In a stream, new mixture.  So, that's a kind of issue thought by CGP.  But from the JGP perspective, we have the characters like which are called simplified or traditional, in Chinese, and we have the same characters, but from the viewpoint of Chinese words, we don't differentiate them with a simplified or traditional.

And all of the combination can be used in Japanese.  So maybe, it's more difficult for Japanese to reduce the number, because all the combination is allowed.  So…

UNKNOWN SPEAKER: For CGP, yes.  Maybe Chinese people can come, CGP members can explain that.

UNKNOWN SPEAKER: So, let me cut in.  Maybe we can take more discussion after the presentation, but we're slightly running late.  So, let's take this forward and move to the next presentation.  We have Professor Kim, who is going to be presenting on behalf of the Korean Generation Panel.

KIM KYONGSOK: This is Kim Kyongsok from Korea. I'm KGP chair. And I want to make sure that okay, it is updated correctly. Thank you. I give the introduction, and the list of [inaudible] characters. In case of Korean RGR, both Han and Chinese characters will be included. Review of K and C variant groups, and then the latest KGR for them, [inaudible] point seven, and something else. Okay, next please.

Characters included for KRD, are both [inaudible] syllables and [inaudible] characters. And the latest one is [four zero point seven], take it much [inaudible] this year. And it has 11 K hundred syllables, and 47 5800 characters. And there are 152 variant groups. In January 2017, the naming [inaudible] of Korea Internet Governance Alliance, abbreviated as KIGA, and its homepage is shown, formally created a working group for allowing the second level and the [inaudible] or dot [inaudible].

And the working group started working in the past, [inaudible] was not allowed and [inaudible]. Next please. The [inaudible] of 11 K hundred syllables are shown. And a list of [inaudible] characters for KGR more than [inaudible] point seven. It is composed of, it is union of two sets, one is KSX one is [inaudible] and [inaudible] in [inaudible].

And in total, it makes 47 58. Next please. And CNK has some conflict in variant groups, and you made coordination and the

result is very good.  And there were about 3,000 variant groups in CLGR, and we don't need to analyze all of them, since Korean characters are not much included in the variant groups.

And in the middle, we can see that there were 168 Chinese variant groups where there are two or more K characters.  K character is a character belonging to KLGR.  If there are no K characters, or just one K character in Chinese variant group, then Korea and China need not coordinate.

And as of February 23rd, there were no more conflict in variant groups between KLGR and [inaudible] GR.  It was discussed and finalized in Beijing.  Next, please.  KGP and CGP coordinate to resolve conflicts of variants groups, and that here are three possible scenarios in resolving conflicts.

Scenario one is Chinese variant group is kept without any modification.  In other words, KGP accepts [inaudible] without any modification.  For example, Korea regards [inaudible] as independent, however, CDP regards those two as variants, then the result is, regards those two characters as variants.

In scenario two, Chinese variant group is fully straight.  In other words, no two K characters are included in a variant group.  They are two examples.  One is C1 and C2 are regarded as independent in KGP.  And [inaudible] regards those two as variants, and the result is making them as independent

characters. [Inaudible] that we have three characters in a variant group, and K regards those three as independent, and C regards those three as variants.

The majority is making them all three, making them all three characters independent. Next please. And scenario three is C variant group. It's partially split, usually one variant character is split from the variant group, and that character becomes an independent character. But a new variant group still contains two or more K characters. Let's see an example.

There are C4, C5, and C6. C regards those three as variants, however, K regards only two of them as variants, C4 and C5. C6 as independent. The result is that making C6 independent regarding C4 and C5 as variants. Here, as you can see, still there is variant group composed of C4 and C5. It is not fully separate. In other words, partially script. Next, please.

And there are lots of numbers. Three or four variant groups, which contain two or more K characters. Korea originally had 46 variant groups, and there is no conflict at all. Now, the remaining number is 258, there is conflict between C and K. And for 258 conflicting groups, [inaudible].

One 10 C variant groups, kept without any modification. There is scenario one case. And 12, C variant group partially split, the

scenario of 3 Ks, and 136 Chinese variant group fully split. Next, please.

As of February 23rd in Beijing, Korea LGR for the [inaudible] points seven [inaudible], contained the 49 19 characters. And a few days later, in March 3rd, Korea finalized KAGR for the [inaudible] point seven, we reduced 61 characters and the remaining number of characters, 47 58, and accordingly the variant group was reduced from, not shown here, but 168 to 152, it is printed somewhere else.

Still, there is no conflict in variant group 15C and A. C and K. Next, please. Okay, it is history. Next, next. Next. Next. It shows the competition of K characters. In the middle, there is intersection of KSX 1,001 and II [inaudible], and in the left hand, you can see 15 characters, which appear only in KSX 1,001. And 138 characters which only appear in II [inaudible].

Thank you.


UNKNOWN SPEAKER:          Okay. So, any questions? Yeah. Mark.


MARK:                    Thank you. Mark [inaudible]. The usual process for doing this work is, people define their repertoire first, and then the

variance, right?  That's, I guess, what you did.  And then Korean and Chinese GP work it out in the variance sets to agree as you presented.  So, my question or, verifying my understanding, that also means, given that you don't want to restart the variant work, that both the repertoire for Korean and Chinese are frozen and will not change, because then you will have to partially reduce the variance sets, right?

So, are you confirming that for the Korean and Chinese, or at least from the Korean side of the table, your repertoire is very stable and frozen.

UNKNOWN SPEAKER:      Thank you.  In case of KGR, we don't have intention to increase the size, at this point.  Actually, we decrease the [inaudible] characters.  It was modified based on the comment from IP.  So, in the future, we will not increase the size of KLGR repertoire, I mean.  Thank you.

UNKNOWN SPEAKER:      We do have another presentation, but I do not see the presenters from, representatives from Thai generation panel, which means we would have another couple of minutes in this session, to take any more questions from either the integration

panel, or Chinese, Japanese, and Korean generation panel members.

So, are there any questions from anybody on any of the presentations which have been made so far?

Yes, please, Matt.

MATT: Matt [inaudible], IS. What is the time plan for integration of Chinese, Japanese, and Korean?

UNKNOWN SPEAKER: So, [inaudible], maybe you want to respond to that?

UNKNOWN SPEAKER: Yes. Maybe we are implicitly requested to be done in this year, but for the timeline designing, we will meet tomorrow, we means C, and J, and K, will meet tomorrow morning to set the timeline. So, I cannot say it now. Thank you.

UNKNOWN SPEAKER: Any more questions? There is not online, so thank you all very much for attending the session. And hopefully, we'll see many of you again in the afternoon for the RDN program update session. Thank you very much.

**[END OF TRANSCRIPTION]**