



# Label Generation Ruleset Process

## Working Group Meeting, Thursday, 30 August 2012

### Table of Contents

Roll Call.....	1
Discussion of draft document approach:.....	1
Issues that we have not discussed yet:.....	9
Presentation on P1: Label Generation Ruleset Tool.....	12
Status of resulting labels.....	12
Examples of Script Mixing.....	17
Section 4 of the updated document.....	19
Appendixes.....	20
Other stuff.....	20
Review of meeting and going forward.....	20
Other topics suggested instead.....	21

### Roll Call

Remote participation: Raymond Doctor, Linlin Zhou, Rinalia Abdul Rahim, Will Shorter, Alireza Saleh, Syed Iftikhar Shah

In the room: Kim Davies, Joseph Yee, Steve Sheng, James Seng, Neha Gupta, Akshat Joshi, Sarmad Hussein, Chris Dillon, Andrew Sullivan, Dennis Jennings, Yoshiro Yoneya, Edmon Chung, Daniel Kalchev, Panagiotis Papaspiliopoulos, Zhang Zhoucai, Fahd Batayneh, Michael Everson, Mirjana Tasic, Dennis Chang, Francisco Arias, Nicoleta Munteanu.

Dennis Jennings: encouraged participants to articulate as precisely as possible to get the most out of the contributions.

Dennis Chang: reminded all that a photographer will be here for 1.5 hours and at some point we will take a group photo.

Dennis Jennings: Today's agenda will go through the proposed panels and how they will work. Andrew produced a new document. We will go through new document.

### Discussion of draft document approach:

Andrew Sullivan: Warned that agenda will not be followed rigorously. The comments from yesterday led to the belief that major changes in section B, specifically expansion on details. Hopes to spend time today to discuss how this expansion should happen. Believes there are gaps in the text where things are not obvious to the reader. Expects to spend on a number of details in Section B. Encouraged participants to send text to the list where specific changes need to be made. Going through procedure as laid out in the document, and if we get to sections C or D, that would be good but not expected.

No one so far has indicated that the two-way approach is not appropriate: a number of primary panels chartered to deal with a script, and one secondary panel to deal with the result.

Asmus Freytag: wants to push the definition of the scope of the primary panel. The primary panel to be charged with a piece of the repertoire. This allows recursiveness. If a repertoire has a piece that is difficult for one panel to deal with, they can assign a sub panel. The sharing of the repertoire is handled by the primary panel. Offered to help Andrew with the details to provide clarity on how that would work.

Sarmad Hussein: two-level is ok, but perhaps there should be public comment after each stage. Suggested that the primary panel takes first pass, put it out for public comment, then have it sent to secondary panel. If there are changes, they are shared with community via public comment. Doesn't change the charter, just opens it more to feedback.

Iftikhar Shah: Supported Dr. Sarmad Hussain's proposal that to get consent from public community it is necessary that public approval is may be obtained.

Michael Everson: Commented that with public comment, one cannot guarantee that everyone will "approve".

Francisco Arias: missing how we are going to define the community or character repertoire? Is it by language, script, something else?

Andrew Sullivan: this is deliberately vague in the draft because different circumstances will call for different answers. Personal feeling is that all Indic or Brahmi derived scripts will be handled by a big panel that will need to break itself into smaller panels. Latin is another example, get various linguistic communities together but they may not need to talk to closely to Spanish users for example. This may not be true for other linguistic communities. Greek will have to cooperate with Cyrillic and Latin but the initial pass will have to come from the Greek community. Different linguistic communities and different scripts have conditions that will need to work it out case by case.

Edmon Chung: general approach and structure is good. Questions about how you envision if there is any overlap between primary and secondary panels? do you envision reuse of primary panels? Constituents of the panels, Edmon said he does not necessarily agree.

Andrew Sullivan: on the overlap question, absolutely not. The idea is that they are entirely different. Impetus for this is that there are gaps in the integrated issues report. Realized that the secondary panel has to be completely different if they are going to be mainly checking on the work of the primary panel. If the members are involved all along, they would not have the same diligence when checking the work of the primary panel.

On the reuse of primary panels, that depends on the nature of the problem. Some scripts are basically straightforward. Once done, there is nothing more to do unless a case of a Unicode change. New uses of Latin will need a reuse.

Asmus Freytag: this problem is similar to creating a constitution. Not writing an operations manual but a very clear requirements on the panels. The primary panel is the consolidator of the input. The suggested model is to divide the work into a top-level panel that may divide the work into lower level panels. The expertise is different. At the lower level or sub panels, the expertise is the writing system. Consolidating that into a proposed repertoire. Interaction between panels: once the proposal has been submitted, there would be interaction between panels where the secondary can check back with the primary on understanding of issues. Consensus, at some panel the secondary panel has veto power. On the other hand, the primary hand can hold things forever by refusing to decide on changes proposed by the secondary panel. There is a nice balance of power because the primary is being pushed by its community while the secondary balances the power.

Andrew Sullivan: the way this suggestion works right now, all the secondary panels are contractors working for ICANN. There is a way to influence by getting ICANN to fire the panelists.

Asmus Freytag: meant that if you have the pressure group to save X language, pressure can happen and a primary panel has a much weaker point. Should explicitly, as each of the major regions are assembled, need to have the straightforward (easy) repertoires. The shell that calls an expert as needed. That could be a standing panel for future input from language experts. If you have primary expertise, you should be on the secondary panel instead of a revolving door primary panel.

Francisco Arias: on implementability, concern is how can we define that someone can speak for a certain community. How can we say that this is the qualified person, and what if more than one party says they represent the same community, how is ICANN to decide.

Dennis Jennings: Ask them to sort it out

Francisco Arias: but still how do we decide who represents a community

Asmus Freytag: once you go on the hunt for a repertoire, you can look at the Unicode. The proposal can go a step further of defining the panels. For example:

Greek/Cyrillic/Latin, Brahmi, etc. That way when people come to the panel, they apply for a certain slot.

Andrew Sullivan: looks forward to chunks of Unicode that are nicely divided up. Last summer we had the six study groups, one of the criticisms we received is that significant users were not represented on the case studies. Examples were the Arabic and Cyrillic studies. The criticism is how can we claim that this is a serious study if major communities were not included.

Michael Everson: going back to before, disagrees with Asmus that people who know a lot about Unicode and context of history, that they be in the second panel. Thinks there is a strong argument for such types of people to be involved in both types of panels, not necessarily having voting power. Example, someone like Michael might have the history, he can be there to advise but not necessarily to vote. Same argument for the secondary panel, someone to bridge the second group to explain why certain things would be there but not necessarily having a vote. Keeping the two panels entirely separate from each other may not be entirely accurate. The second panel is the checker and to make sure the principle of conservatism is applied.

Andrew Sullivan: it sounds like what you are arguing for is a Unicode concierge that goes around between the panels making sure they have the stuff they need to decide.

Michael: one or more people to function in a concierge fashion with different expertise such as DNS expertise, Unicode expertise, maybe others.

Dennis Jennings: voting was mentioned, need to explore that further.

Doc: what happens if the experts in the two panels do not agree?

Andrew Sullivan: short answer is nothing happens. These panels have to agree, if they don't the answer is always no. Partly because a NO could turn into a YES later but not vice versa.

Dennis Jennings: that invites the question of a risk analysis.

Michael Everson: "does that allow primary to say we cannot accept this but could live with that?" situation

Andrew Sullivan: the primary panel will come up with a wide set of what they want, the secondary panel will define what cannot happen from the primary panel's picks. There will be some "theatre" of back and forth between the panels.

Francisco Arias when there is a proposed modification to make changes to the LGR, that's when it makes sense to do public comment.

Edmon Chung: this is where openness and transparency comes into play. Does not agree with where public comment should take place. How do we ensure that when a panel is formed, the community is willing to stand by it? The other going to Asmus's suggestion, it may be a good suggestion, but it would be better for the community to self form instead of ICANN dictating the formation of the panels. Our experience at ICANN tells us that if you don't cater to "lobbyists", the process will not go as smoothly and will cost us pain. Would rather to get the "lobbyists" to participate than exclude them.

Dennis Jennings: agreed that work on boundaries is necessary.

Asmus Freytag One thing that needs to be made very clear is the iterative nature of the process between primary and secondary panels. As such, if there was someone who acted as an advisor, they may not need to participate in secondary. The other issue is that this process is absolutely additive. Another issue is how much of the language community is participatory, how much of the language is written, online, etc. This affects their participation. With the respect to larger scripts, there are certain minority communities that may or may not need to be included. The constitution we are writing will have to say that some languages or communities will have to be considered and added later, it is not realistic to have everyone ready to go. As the high urgency communities are known, it might be helpful for the document to say there will be panels for the known scripts such as Arabic, Cyrillic/Greek/Latin, etc. When the slots exist, there will be a chance for the community to participation. Whereas for the less urgent scripts, there is a shell panel that accumulates data about the need for a new panel to be formed when there is a critical mass for it is there.

Andrew Sullivan: have a question: the fundamentally additive process just discussed seems to cause a rule to change or something to fall out that you'd never want that to happen. The current draft says that if there is a rule change, the procedure is broken because it does not provide the conservatism guarantees.

Sarmad Hussein: initially Andrew started with saying the IDN Variant Issues case studies had some cases where representation was missing. Latin should have been mentioned in those if it was not. Other questions:

1) what is the size of panels we are envisioning? 2) agrees with Edmon on an open call to participate in the panel vs. inviting people. 3) inviting public feedback earlier than later if possible. 4) when something is sent from primary to secondary panel, would the secondary panel evaluate at code point level or the entire proposal?

Andrew Sullivan: to his understanding, the secondary panel is a total rejection or total acceptance. But the secondary panel will also have a feedback loop to say that if this part of the proposal is removed, it will be accepted.

Sarmad Hussein: that approach seems risky. Thinking of the debate on ZWNJ in the Arabic community.

Dennis Jennings: it's actually less risky for the secondary panel to say "NO, but..." rather than making a point by point.

Steve Sheng: it seems that at some point, some compromises have to be made. We've seen some of that with the Arabic community. The process design should encourage communities to work together. The rules should create this incentive.

The other issue is that the secondary panel seems like an ultimate judge. The rules for this panel should be very clearly defined. Anything less will create a lot of criticism and political pressure.

A third point is that we need to set expectations for the community.

Rinalia: wanted to echo Edmon's call for openness and transparency. Very important for the panels to publish their work, agendas and meeting minutes.

Questions: 1) as awareness develops in the community, would it be possible for the panels to grow? 2) Clarification about the bridge between the primary and secondary panels. 3) section 4.2 Expertise lists expertise of the composition of the panel: "A maximal setting places the entire burden for development on experts in the subject area. Experts would be needed in all the relevant topics, including at least Unicode, software internationalization, IDNA, and DNS; for all of these, expertise in both protocols and operations is required." Is this the proposed expertise to compose the panel?

Andrew Sullivan: don't see a worry for that in principle but worries about the panels getting too large. The Latin case for example invites this because of its wide use. The danger is spending time on organizational work and not enough time to create actual proposal. Panel management is the concern.

On the second question: the role of the bridge between the two panels is that those would be subject matter experts making primary and secondary panels. The SMEs on the primary panels function as a kind of bridge, but more concerned about resources, there are a small number of people who get how Unicode works. If we don't have a way of injecting this expertise (position) on the primary panel.

On the third question, the primary panels have a low expertise setting. But the secondary panel should be experts that really understand the relevant technical issues. If we cannot cover all of the expertise listed on the secondary panel, we have a problem.

Rinalia asked for this to be clearly defined because it is not coming through clearly.

Joseph Yee: hears concern about 99% ok but the 1% case of blockage. Can there be a conditional case? Another comment is that he likes Michael's idea about advisers but worries about negotiation too much between the primary and secondary panels. It should be someone who has a high level expertise in the community that can raise issues whether they are accepted or rejected.

Panagiotis Pappasplipoulos: How can YES become NO? Agrees with Sarmad's comment about public comments. What is the process for changing a YES back to NO?

Andrew Sullivan: imagines a less formal setting. That primary panel reaches agreement, then run their proposal by the secondary panel in a collaborative mode where the primary panel is given indication of what in the proposal will cause rejection. This will give the primary panel the chance to edit the proposal and send it back.

About how a YES changes to a NO, the process will have failed if something is in the repertoire that needs to change later.

Panagiotis Pappasliopoulos and Daniel Kalchev: will there be an appeal or decision reversing process to change the decision of a panel that may have been made by mistake. It would be too late to do this after the secondary panel has made a decision.

Andrew Sullivan: in such a case where a mistake is obviously has been made. We just have to rely on the secondary panel to be able to recognize the error. This is why the public comment is helpful. This is why the secondary panel of experts is a good idea with no lobbyists. Somebody needs to be the final decision maker, they have to explain their rationale for decisions made. They have to be able to defend themselves.

Acknowledged that the IDN ccTLD Fast Track was designed with no appeal built into it and that has caused some criticism.

At the end of the day, knowledge grows and in the light of new experience things could change.

Asmus Freytag a NO vote has to have a detailed with rationale. It cannot be a No only vote.

Daniel Kalchev: the world outside ICANN is much bigger. There may be experts outside of the ICANN realm that have more knowledge. If the rules are already implemented, it is too late to appeal. There may need to be an experimentation period or a way to test the rules.

Dennis Jennings: this is an interesting point that needs to be thought through.

<<Morning Coffee Break>>

Resuming discussion about coffee break

Michael Everson: We should have discussion on membership of panels. Attending a panel is not necessarily being a member of it. Thinks that it would be useful for the first stage panel to have expertise that would be key for the second stage panel to draw on their knowledge.

Joseph Yee: transparency is important.

Michael Everson: related to what Asmus said, wondered if the general public consultation. In ISO, they tried to acknowledge comments. Depending on volume may not be feasible, but an analysis of the comments should be shared.

Asmus Freytag wanted to take a step back and from his experience on Unicode experience which is by nature an additive process and once things are added, they cannot be taken away. Has hard-learned experience where you always act on limited knowledge and may have to reconsider things later. Also the Unicode does not have the same luxury with regards to the conservatism principle. The Unicode tries to make its process as thorough as possible, but despite that it still has less than perfect output. Desiring a perfect output cannot be guaranteed. There may be linguistic communities that emerge later that will find itself locked by current rules. A remote possibility but there nonetheless.

Also, one more point, which panels to constitute: 1) scripts and repertoires for which there is known interest and 2) scripts and repertoires for which there is NO known interest. For these where there is demand, it would be easier to start where demand exists.

AS: what constitutes demand? Someone who thinks this is all absurd might come with a claim to represent a new language that is extremely obscure yet there are people who need it? Do we need or want to be in the position to say that creating rules for this is too hard and we don't want to do it.

AF: we can make one exclusion rule, if there is no money on the table to apply for a TLD in that repertoire, there would be no demand for that script. This would mean there are large sections of the Unicode repertoire that would never be evaluated but thinks it is a good thing not to randomly create a list of what to evaluate.

AS: wanted to clarify why what Asmus said differs from the current draft.

AF: differs in the level of comprehensiveness.

AS: description of extremes of parameters, people have interpreted as somehow constraining the way things are going to proceed. The way the document is now laid out seems to be missed by readers

DJ: In the current document, the two extremes side by side and what is in the middle

AF: suggested a more drastic change, to leave the theoretical section until the end and put the proposal at the beginning. The theory section is to say: "what makes us feel this proposal is good".

Sarmad Hussein: IDN ccTLD applications already delegated, and new ones that are in the new gTLD program. Whenever a new script is applied for, it kick starts the work for that new script. The question is when would that work start? When the request comes in or when the application is approved?

AS: seems to me when the application is made, the demand for that script is established.

Sarmad Hussein: does not have a particular preference but this may be something to get added on the procedural questions.

Edmon Chung: Agrees with Asmus's suggestion and happy that Andrew agrees. Should start with those that demonstrated demand and moderate it that while it may not be proper for us to create the panels. We should do outreach and proactively to see communities' needs. Be proactive rather than reactive.

End of discussion on approach.

### **Issues that we have not discussed yet:**

AS: which substitutions are going to be permitted. Cases for 1:1 and 1:M. On the call from 22 August, the envisioned variant rules are M:M. The danger with such an approach is that you quickly get on the variant discussion based on semantic basis.

Sarmad Hussein: wants to add Tibetan letters. There may be other scripts that require contextual rules.

FA: if remembering correctly, the case in Devanagari, the issue was string similarity not exchangeable code points.

Akshat: should be focused on syllable or akshar. In Unicode, for some conceptual characters, Unicode started with one character then changed to a new more appropriate character. The example is the nukta characters. Another example is the eye-lash rah. The mechanism of inputting this character needs ZWJ, later on, the inputting method was suggested to be changed to be without a joiner. Referred to chapter 9 of the Unicode. To Francisco's question, yes the cases in Devanagari are visual similarity, but the cases should be the basic building blocks.

DJ: what is a basic number of characters in a building block?

Akshat: could be 9-10 characters considering the worst-case but in general, 3,4, or 5.

Sarmad Hussein: for Lao, it is 9-10 characters as well.

Asmus Freytag needs to be guidance. There are certain kinds of features in Indic script for sequences. As knowledge has grown, changes were made.

Akshat, demonstrated on the whiteboard some of the most complex syllabic examples. Gave the example of Kitab.

Another case of two cases of eye-lash characters where there are different ways to encode the syllables which causes ambiguity.

FA: as of now, is this the only case?

Akshat: that is currently true but maybe change.

FA: trying to get the size of the problem, it seems there are not many cases. Is there a definition to be used about the basic building blocks?

Sarmad Hussein: there are two different issues here concerning the example given and the document. 1) whether a non-letter should be allowed: 0940 Halanta is not a character, it is a mark needed to write Hindi. and 2) whether context rules should be allowed. When you originally look at it you see a vowel then a consonant. If someone for whatever reason types it visually, vowel then consonant, it should be a contextual rule. So perhaps there needs to be some contextual rules to manage strings.

AS: Three problems: 1) checked this last year and at that point the normalization of this does not result in the same code points. NFC does not make this problem go away. 2) while it is true one of these characters does not have a letter property, that is why the IAB document adopted the letter principle. And finally 3) why M:M is not perfectly acceptable, it quickly leads to color and colour being variants of each other. Put less positively, someone will attempt to use this justification for why they should get this variant.

Raymond Doctor: wanted to repeat what was said yesterday: 1) That the akshar as demonstrated by Akshat is the basic building block of all scripts derived from Brahmi 2) that the notion of the syllable/akshar is inherent to Indian scripts and old grammars attest the same. And 3) that since the Akshar constitutes a context-bound writing system, it is requested that such a system be incorporated in the draft proposal.

Asmus Freytag: thinks we can go beyond Brahmi derived scripts. There are other scripts where that may be the case. How do you write rules that prevent structure nonsense from being registered? Just because we don't have that in standard alphabetic scripts, doesn't mean it does not exist. This leads to the next point, somewhere in the document there is a section that once the rules exist certain label validation rules can be automated. A secondary panel should in fact be one that writes algorithms. The secondary panel can go a step further and create an algorithm. With an appeals mechanism.

Can conceive for the moment that saying that certain features of the algorithm saying that certain scripts do not require the M:M rule but others do. Being able to create that rule on semantic considerations only. Want syllable substitution rules.

Wants to open the possibility for different approaches instead of one size fits all.

Edmon Chung: agreed with Asmus. Any reason why this must be so definitive? Why can't the structure of primary and secondary panels solve these issues and 3) we already have an understanding that for well-defined cases such as Japanese, otherwise script mixing is not allowed. If this whole process allows for something like "color" and

“colour” to go through the whole process, maybe we should believe in the structure and let it be.

AS: it’s true this is a procedure that will rely on human judgment. But within that, we need to have well understood rules because if we do not, every one of these decisions will become a political decision. It is the belief of this program that the ICANN Board for example does not want to decide on a one-by-one case and instead a predictable, repeatable process.

The reason for the definitive statement in the document is the simplicity rule.

Asmus Freytag: the definiteness is necessary but would like to see it more explicitly explained. Giving the secondary panel some flexibility to evaluate the risks and providing restrictions. This allows discussion between primary and secondary on the issue.

Michael Everson: presumably we will have a table of equivalents, why is it a problem to have a table that expresses M:M characters? A table of excepted variants would define these rules.

Akshat Joshi: Regarding ZWNJ, in some cases the visual difference is seen and not in others. But in the basic set of rules, these complexities are not there and we can easily predict the outcome.

AS: The way the document is currently written: if you understand enough about Unicode and IDNA, the set of rules should be reasonably comprehensible without understanding the script. But the idea that anyone reasonably competent will understand the rules. The more exceptions we have, the more we cannot meet this rule. Essentially, a rule that is peculiar to the script itself, and only knowing the script or language will allow you to understand it. That maybe a justifiable position.

AF: one basic rule is that a well-formed combining sequence must start with a base character and may not start with a combining mark.

AS: to answer Michael Everson’s question, one answer is to produce tables. The point is to permit useful mnemonics in the root zone.

Context rules: in IDNA, four categories: PVALD, disallowed, ContextO and ContextJ: ContextJ says joiner character is permitted only after a character that has the property virama. If that is not the case, the character is disallowed.

AF: playing the devil’s advocate, what if that letter was not ZWNJ and instead was a more used character, can you contemplate a context rule for that case or not?

AS: as a general matter, they make the rules much more complicated. We are talking about the root zone. The goal is to allow useful mnemonics. Does this cost more than it’s worth to the global community?

AF: what if these characters appear so frequently, locking them would disadvantage a large community. Sarmad has an example to demonstrate this case.

Sarmad Hussein: English or Latin: any strings let's say "first" cannot be inputted because "st" cannot be inputted. Is that acceptable?

Akshat Joshi: yes currently there is no example, previously the nukta character fell in that category. Generally such cases are dealt with the help of Unicode normalization. We need a way to deal with this if it ever comes up again.

Doc: sees the argument is around eye-lash rah. It was an international company that created the rules for the eye-lash rah. Another example is the Akshar. ISCII provides for rules to deal with this correctly. Unluckily, word processing companies still do not process correctly. Only Nepali requires ZWNJ.

<<Lunch Break>>

### **Presentation on P1: Label Generation Ruleset Tool**

See presentation from Kim Davies on:

<https://community.icann.org/display/VIP/P1+Label+Generation+Ruleset+Tool>

Historically ICANN is contractually obliged to publish their IDN tables in the IANA repository. ccTLDs are encouraged to do it.

There is currently no consistent format in which the IDN tables must be represented. Having a universal format helps us to do comparison and analysis between formats.

Kim reviewed the goals and non-goals listed in the presentation.

Current status of the project: draft has been published. Draft -02 pushed recently.

Seeking more feedback.

### **Status of resulting labels**

Andrew Sullivan: Question of the status of resulting labels, there is little discussion in the draft about what happens with the resulting labels. In the issues report, we identified different states: Allocation, delegation, blocking, withhold. There's not much discussion for how those things happen. There is not much discussion on the blocking because blocking is defined as never allowed.

A related question is whether we want the rules to specify this. Whether we generate the variants and it is up to the applicants to say what they want done with it.

Asmus Freytag: we just decided earlier to block structurally non-well formed. Would that be only as related to variants because what we defined is non-variants?

Edmon Chung: any well-formed policy needs to take the resulting labels, certain possible status should be assigned: automatically delegated, optionally delegated, withheld or cannot be delegated.

Don't know if we need to cover this in the present document.

Andrew Sullivan: One way to think about it is that all the discretion lies with the applicant, and there is a policy that says that if you have such and such codepoint, it generates these variant(s) and the applicant decides whether or not to turn on the variants and that automatically restricts what the panel can do.

Edmon Chung if that is the starting premise, he would disagree, that would not be a good enough framework. The Optionally delegated is what Andrew raised, Edmon believes there are two other options: must be and must not be delegated. Those must be covered.

Andrew Sullivan not addressing those two other options reduces complexity.

Edmon Chung looking back at the study team reports, there are requirements for variants, and a particular premise that there are certain variants that must be delegated for certain domains. To take away that possibility would be to unravel the discussion of variants.

Sarmad: Minimally we get a yes/no answer against a label if you are querying for a label. Additional is a list of variants, then a distinction about the hierarchy: primary vs. fundamental. The requirements seen from previous phases is that they are not the same across all scripts. For example, in Chinese, there is a fundamental label, but that is not the case in other scripts. In Arabic, there are two different fundamental labels, one to be typed on a keyboard from Saudi Arabia vs. one to be typed on a keyboard from Iran. If a Saudi Arabia person is applying, they would be the primary label vs. if the person from Iran is applying, they would want their label to be the primary. 1) how many categories to do we want, and 2) is there a mechanism that should be made part of the label generation process that something is primary or secondary and 3) is there something coherent across all scripts?

Asmus Freytag in the case of eye-lash rah, we would not want the applicant to decide. This is a matter of semantic. Sees how it would be simpler for panels not to decide and leaving it to the applicants to decide. Cannot see a case where they applicant would be told you must have both enabled, demand must be the deciding factor.

Edmon Chung It is not a matter of technical consideration but specifically on policy. On a policy level, that is what is required, not what the applicant requires.

Francisco Arias would like to have the flexibility for blocking variants. On visual similarity, that is a blocking matter. Regarding the "must have", remembers a discussion from the Chinese study, it would be helpful to have an example for why.

Edmon Chung it would be in compliance with CDNC policy.

Francisco Arias in the Chinese community, variants are considered the same name. Understands that technically they are two different strings but this is where policy comes into play. If an applicant is only interested in serving one, that is not an acceptable policy at the Chinese community level.

Dennis: what about an applicant that does not have that requirement such as an Irish applicant wanting to serve a small community in Ireland. Why should they be forced to have both?

Edmon Chung because it affects a broader community, bigger than the Chinese community only in Ireland.

Joseph Yee: Unicode may only care about X but there are other considerations that need to be taken into account.

Andrew Sullivan It sounds that there is strong support to specify the disposition in labels. Back to Sarmad's point for naming a primary, the draft that all of the rules are completely symmetrical and in that sense none of the is primary. If we want asymmetric rules, it makes the process more complicated. Personal feeling is not to allow a symmetry but open to suggestions.

Sarmad: allowing asymmetry allows for fewer strings in the root and that is a good thing.

Andrew Sullivan but it also forces the application to say that depending on how the app comes in, certain states would be automatically decided.

Asmus Freytag If we are considering the exclusion of widely used mark from a particular writing system, that would severely hamper certain mnemonics, the extreme example is to say that CJK variants are not going to be applicable in the root. If this position is taken, that would automatically exclude variants. It's an interesting concept to consider.

Andrew Sullivan symmetric rules are understandable, it is straight forward. But deciding its status depending on how it is submitted makes it much more complicated.

Asmus Freytag: would a Latin a block Cyrillic a

Andrew Sullivan In a symmetric relationship, a Latin a would block a Cyrillic a and vice versa. In an asymmetric rule, a Latin a would not block a Cyrillic a and vice versa.

Daniel Kalchev: comment made earlier about understanding of language and scripts. From somebody who is not familiar with Chinese, they will not be interested in how Chinese variants will work. The only interest is in the work on the tables are done. When the tables are done, you don't even have to know Chinese either. Why put so much focus on the knowledge of each language and variants? As long as we have the tools such as the tables?

Andrew Sullivan the draft does not say that everyone has to have knowledge of all languages. The principle of simplicity and predictability, the idea is that for public zones, what you need is something comprehensible for people that don't know the script or language, in order to know why the rules would need to be structured in one way or another, you would need to be able to speak the language.

Daniel Kalchev: but as an implementer, all you need to know is how to interpret the rules.

Andrew Sullivan agrees that not every implementer needs to understand the details but still a certain amount of background is still needed to decide what applies to you and what doesn't.

Daniel Kalchev: disagrees, the implementer does not have to have any knowledge of the result. Make the proposal clearly define the output of the primary panel and the work of the secondary panel.

James Seng: where in the document is the discussion on symmetry.

Andrew Sullivan thought this was stated, but apparently not. It's only in Andrew's head at this time and needs to be added.

James Seng: if added, I will oppose it. There is no symmetry in Chinese variants. James gave an example of a presentation from his desktop.

Edmon Chung this is symmetric under the CDNC table.

Steve Sheng: wanted to see if there is consensus in the room. One of the problems variants are trying to address is security issues that may arise. Also thinks that for the Chinese, the issue is not to confuse users by delegating to different applicants. Is it ok to say that by default, they are allocated to the same applicant but withheld.

Edmon Chung labels generated through the LGR would have different statuses. What are you asking?

Steve Sheng: trying to limit delegated variants. Thinking from a usability perspective.

Edmon Chung that's why they have the concept of a preferred variant. Don't want to have too many but there are certain ones that are the crux of what has to exist. The best way is to look at the table, there are certain cases where there is more than one preferred variant but those are rare.

Francisco Arias: regarding the must be delegated category, perhaps of what would be accomplished because of mirroring not being an option, given that, would that accomplish what needs to be accomplished?

Edmon Chung the intent is that they do not mirror or behave the same. That is not the expected user behavior. The user expect them to come from the same applicant.

Francisco: on the policy level, why does someone have to have the variants if they are not identical. I completely see that if the variant is ever allocated, they would need to be allocated to the same applicant.

Edmon Chung if you are trying to talk about content, we should not go down that path.

Asmus Freytag why can't the variant be withheld.

Edmon Chung It is a body of work the Chinese community has. The statistics from China TLD that has been delegated, 20% of queries end up in the variant. To Chinese, they are basically the same name. It requires a big body of work to substantiate that claim. This seems to be fairly unique for Chinese situation.

James Seng: 70-80% of Chinese TLDs have variants. From the end-user experience, I may see traditional and simplified but my keyboard may only allow me to type simplified, so this is to improve the user experience.

Dennis Jennings: as a hypothetical entrepreneur in Dublin wanting to serve the Chinese community in Dublin, why should I care about the traditional if my clientele is all simplified.

James Seng: Nothing forces you to use them, at best it is a lame zone not to use. You will get traditional queries whether you like it or not.

Steve Sheng: this is also good from a security point of view.

Asmus Freytag symmetry of rules would require it to be used with similar policy.

Edmon Chung an example is the capital and small letter A. it is the same but knowing this is a good example, the closest to use to try to explain.

Asmus Freytag in the root, you want things predictable. if that is accepted, why not apply this solution universally.

Joseph Yee: Acknowledge this is unique to the Chinese case.

Asmus Freytag still don't see why such a concept would not apply more universally.

James Seng: in this case, there needs to be a definition of variants. In the new gTLD process, they are the visually similar strings. UTR36 and UTR39 addresses this issue.

Asmus Freytag UTR36 and UTR39 do not address this issue.

Andrew Sullivan having determined the ability to say this must be turned on vs withheld vs blocked is what we are trying to get to so this does not need to be resolved right now.

<mid afternoon break>

Resuming after break

### Examples of Script Mixing

Dennis Jennings met during break with Akram Atallah, acting CEO and COO and is pleased to report to the team management's full support for the project.

Michael Everson: still has not heard a definitive case for allowing script mixing. Script property of Unicode code points in the draft is currently allowed. Should not be allowed because the word "example" for example can be spelled in 16 different ways of spelling the same word because the mixing of Cyrillic and Latin is allowed. Everson gets 19 combinations. This is compounded if the word gets longer.

Andrew: the way the proposal reads, the panel does not strictly adhere to script property. What that does, it permits as a simple consequence the Japanese script mixing. It allows inclusion all the things that would be allowed without relying on the script property but the script mixing as explained by Michael Everson is not allowed.

Michael Everson: would like to see simplicity with exceptions. The default should be no script mixing unless a need is established. Some panel has to look at the content of common category and decided what is allowed.

Francisco Arias: reminded the group of the rules of the new gTLD rules to specify no script cases except in known cases such as Japanese.

Andrew: yesterday there was call to permit the hyphen. If the hyphen is allowed then everything in the common has to be included.

Daniel Kalchev: another example "Скoпjе". This looks like it is in two different scripts: Cyrillic script, Macedonian language with common characters in Latin.

Asmus Freytag: the indistinguishables do have usability issues. Both Daniel's and Michael's examples explain this. Gave another example mnemonic "exec". May have a Russian and Latin domain. As an international venture, may want "exec" in Cyrillic and Latin. Could see a registrant that needs both so a user can get somewhere when they type the query in Cyrillic or Latin. If we construct the rules to work for Chinese, if we want to keep the rules consistent, the rules must allow for such case rather than universal blocking.

Daniel Kalchev: another example “3com” people might think 3 is a number but it is a letter in Cyrillic. The 3com corporation may want this solution for their company name but it is a reasonable abbreviation of something.

Francisco Arias: thinks these examples are related to use experience. What does somebody type when they see “exec” on the side of the bus. Reality is if the bus is in Russia they will type it in Cyrillic, if the bus is in NY, they will type it in Latin characters.

Andrew Sullivan: section 4.3 needs more text to respond to the question Michael Everson raised about script mixing or cross script rules. Rather than nailing the whole thing to script property and using exceptions, give a script tag to everything that belongs to the repertoire. This prevents from creating an algorithm with exceptions.

Asmus Freytag: this enforces that all the characters in a label belonging to the same context would get tagged as such. There must be a rule that says the primary repertoire has to have a definition with tags assigned to each code point. The tags are not based on language or script, it is by repertoire. This prevents repertoire mixing.

Daniel Kalchev: going back to the tags. How would you propose the characters in the example given “Скoпje”.

Asmus Freytag: This is a case where a registrant might want to register everything is in Cyrillic. Cyrillic typist cannot type “j”. Moreover, this is not part of the Cyrillic alphabet and would not imagine that it is in Cyrillic.

Andrew Sullivan: it may be that the label “Скoпje” may just not be allowed. This would be mixing script.

Daniel Kalchev: if a Macedonian applicant applies for this, this is for Macedonians is their version of Cyrillic but not to everyone else using Cyrillic.

Michael Everson: there are over 400 Cyrillic characters. If the Russians cannot use a character user a Macedonian Character, this is their problem.

Daniel Kalchev: this is the Unicode’s fault but I will stop here with blaming. What we don’t want is for somebody from Russia to apply for this string. How do you restrict a Russian apply for such a string.

Dennis Jennings: why would anyone do that? And if they do, what problems does it lead to?

Francisco Arias: are we missing a “no script or repertoire mixing” in the document? The second point is about Asmus’s example about two visually similar labels cannot be delegated.

Andrew Sullivan: wants to think about it more. Will try out some text and send it to the list.

Edmon Chung: on the similarity, the staff implementation does not allow it but gNSO made it clear that there are cases where visually similar labels might be allowed.

Sarmad Hussein: there are so many characters in Arabic script and anybody can choose a subset and register a domain and whether or not the characters are available on a keyboard is not relevant. It is generally applicable across all scripts and does not see that this issue is specific to Cyrillic.

Daniel Kalchev: It's different because it's not like using different Cyrillic characters. In the example, people from Russia and Bulgaria will always recognize "j" as a Latin Character and not a Cyrillic one. This is about the perception of Cyrillic users that this is not a Cyrillic letter and it is from Latin.

Historically this is a letter from the Cyrillic alphabet. This was created by the Unicode. For those that were mixing Cyrillic and Latin characters and labeling them as Cyrillic or Latin.

Asmus Freytag: acknowledge that certain issues are caused by the history of things but we should not try to solve this particular problem here and we should not try to solve some usability problems the same way. It has been said in this discussion that the root is shared without context, all you can tell about label is only by looking at the displayed form and starting from there. Usability issue that have to do with the internationalization of the root have to be addressed, some are detail issues that can be pushed down. Wants to hold the door open to at least discuss these issues in the panels. Proposing to capture that somehow these issues belong in the final settlement somewhere else.

Michael Everson: In response to Daniel Kalchev, it's not Unicode's fault and looked up the history of when it happened.

Dennis Jennings: did you get enough input from this discussion for the next generation of the document about script mixing?

Andrew Sullivan: Yes.

Michael Everson: If the rule is somehow there cannot be mixing, that would have to be finessed in the text.

#### **Section 4 of the updated document**

Andrew Sullivan: if I missed anything in section 4, please comment and send me text. That is true for the whole document, but specifically section 4 are consequences of things we talked about yesterday.

## Appendixes

Andrew Sullivan: Added these appendixes to address Michael Everson's call for more examples, are these the right headings?

Asmus Freytag: Guidelines for Evaluation Process is something needed. If the result in section 4 is detailed enough, that is ok, if not, needs to be somewhere else.

## Other stuff

Andrew Sullivan: Comment 19 in the document. What do people think of this?

Asmus Freytag; have no problem with looking at the procedure. With increasing knowledge, the procedure after further review may need review. It is not something wrong with the procedure necessarily. Mistakes will be made, there is no way of avoiding despite the best efforts. The other thing is that the only consequence cannot be just to review the procedure but also how to minimize the damage, a general risk analysis. Relies on Andrew for impact on DNS.

Edmon Chung: Agrees with Asmus's position that if later down the road a repertoire is not correct, we should put a better one in place and of course we should study the impact such as does a TLD need to be pulled? When we hit this point, it becomes a much bigger discussion. If .UK were to be suggested now, it would not happen. Overall policy should not stop us from improving the policy even if it impacts the existing ones.

Asmus Freytag: if a mistake later is discovered, here are the consequences for grandfathering, etc. these actions would be taken by the Secondary panel. Overtime, the secondary panel would have the responsibility to handle unanticipated issues.

Andrew Sullivan: hears the message about potential failure. But this procedure is designed to say that if I am not sure, then the answer is no. Only the things we are absolutely sure about are the ones we would permit. If the process produces a failure, then the process has to be opened up and analyzed to find out the failure, and new additions to the root would be stopped in the mean time.

Asmus: the first level would be to establish if the error could have been avoided with a change in procedure. If that is established, then everything is put on hold until the procedure is reviewed.

Andrew: the skeptics about expanding the root would use that in their argument to say to not add any more to the root.

Asmus: Judgment based on incomplete information is unavoidable.

## Review of meeting and going forward

Andrew Sullivan: opposed to this agenda item because of the large number of things that have to be changed.

### Other topics suggested instead

Edmon Chung: composition of panels, we touched on it but did not expand, when do we give input on this?

Sarmad Hussein: another topic, how many labels can be activated is another topic we brought up and did not reach closure.

James Seng: there should be a session on security consideration and things done in UTR 36 and 39.

Yoshiro Yoneya: discussed previously how and who defines the mixed repertoire

Michael Everson: made a suggestion about 4 codes that currently exist. The Unicode has not done anything with JPAN. Could suggest it to the Unicode Consortium.

James Seng: do we adopt the ISO definition of script block?

Asmus Freytag: no such thing as script blocks, but they are not a logical division. They are based on the collection of shared typographical styles, not based on usability. The ISO script codes are to answer a different questions such as script in which a document is printed.

James Seng: based on IIR, the definition of script is according to TR24.

Michael Everson: the registrar of ISO 15924 and a suggestion for handling Japanese so we don't even have to say script mixing.

Andrew Sullivan: will propose new text. Likes the idea of using ISO 15924. If people are agreeable, that is what we'll do.

Dennis Jennings: we need a first piece of example text of the issues raised above.

Andrew Sullivan: two follow up of questions 1) on the composition of panels, what did Edmon mean?

Edmon Chung: needs to flush out. For example, not convinced that secondary panel should be paid consultants by ICANN. What was the motivation for that?

Andrew Sullivan: could be staff members, not necessarily consultants. The fundamental purpose is to check work. Should be highly resistant to community pressure. Should be people who are "disinterested." Don't know how it can be done without hiring the work. The secondary panel has to be beyond suspicion of interest.

Daniel Kalchev: can you say non-objective interest?

Andrew Sullivan: even if it completely benign, people with commercial interest have the tendency to open a bigger scope to increase the potential commercial interest.

Daniel: but we still need to have those people to be motivated somehow. The motivation to receive payment is not useful.

Asmus Freytag: we have a professional responsibility toward the goal of making sure that these people are there for the purpose they are hired to do.

Andrew Sullivan: the primary panel is supposed to be interested and advocates and should weight this against interest. The secondary panel should be the world's best experts. We are not going to get the world's experts for free.

Francisco Arian: comment from Rinalia: some experts may not want to be paid by ICANN to maintain their independence.

Edmon Chung: echoed Rinalia's comment. Same argument applies to the ICANN Board, and we can dig up all the arguments for why it is not the case for a specific purpose. While the intent is to protect the root, those are situations where you want to have ICANN policy development involved. Does not want to rule out this possibility at this time.

The other thing is that we talked about a person that floats between the two panels. Maybe an observer or non-voting members.

Francisco Arias: clarification question to Edmon, what kind of registry/registrar implications do you foresee? The LGR is for the top-level.

Edmon Chung: There are certainly implications on registrar and registry communities.

Andrew Sullivan: in the DNS, if you have an active variant, the registrar will have to manage to kinds of input because from the DNS point of view, those are two different domain names.

Daniel Kalchev: the argument of people who are paid to do the job are more independent that those who are not is very weak. Asked if we can go back to the previous question on variants and the example Edmon gave with the Chinese case

Sarmad Hussein: 1) what are the logical states independent of the technological solutions. And 2) how does a TLD move between states? A higher level question is that this this something that this group addresses or is this question handled somewhere else?

Francisco Arias: the states are solved in the IIR: blocked, withheld, allocated, active and mirrored.

Sarmad Hussein: IIR was supposed to raise issues, not solutions. Will this be adopted as the strawman by reference to the IIR?

Dennis Jennings: yes, this can be adopted as a starting point.

Francisco Arias: on the conditions for state changes, that maybe in the scope of the Usability project.

Andrew Sullivan: in at least some cases, the label generation rule will either block, some might be withheld, delegated, etc. This procedure of LGR has something to do with it, but the procedure by which it happens is outside the scope of this project.

Primary and secondary panels: determine the repertoire and the rules for allocation, delegation, etc.

Francisco Arias: agreed that panels would determine the repertoire and label generation rules but the requirement of “must delegate” would be up to user experience considerations.

Joseph Yee: have not yet made the distinctions of all the rules the panel operates under. The rules generate all the permutations: whether that would be decided by human decision or a computer generated algorithm. Has this been decided?

Andrew Sullivan: The answer depends on what the user experience panel comes up with whether or not it is permissible.

Sarmad Hussein: it should be articulated very clearly what the applicant is going to request. What the applicant will generate, what the LGR will generate and what states would be assigned to each of the states.

Francisco Arias: Once the LGR process is there, the applicants make their application and the process depends the variants based on the rules in the LGR.

Edmon Chung: even if we have the LGR, the applicant should list out the variants to show that they know what they are dealing with. If they give us the wrong variant label, that is a good indication they don't know what they are doing.

Dennis Jennings: proposed to bring the meeting to a close. This has been a very good face-to-face meeting. The next one is in Toronto. A huge amount of work to be done between now and then. The revision and version control has to be considered. We will also be talking about this in Toronto publicly. Deadline for first draft for public comment before Toronto is 24 September.

Francisco Arias: conference calls are not in the current plan but if people want it, it can be arranged.

Dennis Jennings: believe a conference call is important, perhaps in two weeks time.

Edmon Chung: agrees with conference call. But what is more important, before publishing for public comment or after?

Andrew Sullivan: happy to have a conference call but comments via email are a lot more helpful. Really wants notes sent in writing, much more helpful that way.

Michael Everson: wants acknowledgement of remarks sent.

Dennis Jennings: when will the next revision be done?

Andrew Sullivan: will be able to give an estimate tomorrow.

Dennis Jennings: we need to define a timeline. Project team will work on one and share it.

<<End of two day meeting>>