CR - IDN Variant TLDs Issues Report and Next Steps
Wednesday, March 14, 2012 – 09:00 to 10:30
ICANN - San Jose, Costa Rica

Dennis Jennings: Good morning everybody. Oops, big echo. That's nice. Thank you. You want me to sing? Sure. It's 9:00 so we'll start. Kurt, would you like to make some introductory remarks?

Kurt Pritz: Very few. To those of you that got up early, thank you very much for coming. I'm surer more people will join as the session goes on because it's very important work and it's generated a lot of interest. What's being presented today are two things. One is the results of your collective work on the IDN Variant Issues Report. It's really a very important, meaningful and significant amount of work. Many community members contributed many, many hours, traveled, contributed facilities, meeting spaces, people's time to putting in the work developing this report.

It's the culmination of six separate Variant Working Groups and then the integration of all that work into this final issues report. So we're going to have a description of the results of that report first, and then we're going to discuss the plan for going forward next. And the issues report did its job; it lays out all the issues before us and they're not low in number and they're not simple. And so the project plan seeks to address each one of the issues raised by the community; the community members that worked on the report.

And so the plan for going forward is very comprehensive. Anticipating some of the comments, the plan for going forward is not only comprehensive, it takes a good bit of time and a good bit of ICANNs or someone else's money – actually all of ICANNs money is someone else's right. So we're interested in getting your feedback on it. If you think the timeline is too long, we're interested in getting your ideas in accelerating it; either now or later. We're marching down

the path continuing to do work. We can continue to talk about how that work is prosecuted and a timeline for doing that.

So thanks in advance for participating in this, and thanks in arrears for all the work you put into this. And it's my honor to present Mr. Dennis Jennings who himself personally has done so much work on this project, calling from the middle of the night in Ireland to the project team on a weekly basis and all days in between. So thanks very much Dennis and thank you to the team.

Dennis Jennings: Thank you very much indeed Kurt; thank you for those words. It's been an honor and a pleasure to have had the opportunity to provide some leadership to this project. But the real kudos go to, as Kurt has just said, to the teams, the community teams who did the work, and to the ICANN staff team who did extraordinary work in very short periods of time to get out the report. Okay, let's bring up the first slide. Oh, I've got the clicker. Thank you. So this is our agenda here.

I suppose I should preface all this by saying that although we're on a path which we all hope will lead to the delegation of IDN Variant TLDs, there is no assumption in all this that that will be the outcome. There is no guarantee that we will be able to address all the issues. So this is just an alert that there's no certainty that we'll come to the end and delegate IDN Variant TLDs. But we do know that there is a huge demand out there and therefore we're doing as much work, the community will do the work, to try and get there as you'll see.

So, we hope to spend about 10 minutes just looking at the overview of the projects and the goals. 25 minutes on the work that's been done and the report that's been issued, and I hope that everybody has read the report. It's difficult, it's head wrecking stuff, but it's important to actually read it, because regrettably this session is not a tutorial on Variants. It's not a tutorial on code points. If that's necessary we will happily organize that, but this is all about presenting the report and getting feedback and then the next steps. Next steps is the third item and then discussion and questions; a lot of time for that.

So, what's the overall; we'll just review this. It's a long standing request from the community to have IDN TLDs and IDN Variant TLDs, and it's the subject of a Board discussion and a Board direction to develop an issues report on the subject, and there's the link there. And the Integrated Issues Report is the outcome of that work.

So let's get down to the details. Naela, can I ask you to take over and to tell us about the work that's been done?

Naela Sarras: Yes, thank you Dennis. Good morning everyone and thank you for coming to this session. We thought we'd start out this morning by giving a little bit of an introduction on what was done in the last year. So as Dennis mentioned, this was a long standing request from the community. We started the work, w officially launched the work at the beginning of the year. I think around March or so of last year. We put out a paper that said "Here's our approach," and our approach was to do six case studies. As a matter of fact it was five case studies of five different scripts and then the community gave us input that we should probably add a sixth ones based on need and the community also specifically helped us refine the Devanagari Case.

So those two steps, the first set of arrows, represent the six case studies that ran from about May until October of last year. And then we went into the integration of the information that we learned from these six case studies, and that was phase two, the Integrated Issues Report. And the outcome was in December, the final document.

The way the six case studies were selected again were based on consultation with the community and we identified Arabic, Chinese, Cyrillic, Devanagari, Greek and Latin. The teams were community teams that were at least about 15 to 20 I believe to each team; some teams had less some teams had more. They had expertise in different areas that we were looking in order to have a comprehensive report. So we had DNS, registry, registrar, linguistic, security, policy, IDNA and expertise on each team.

And the teams worked very hard. We met periodically. We had phone calls. We had, with most of the teams we had face to face meetings. And the outcome of that work was an issues report with issues about Variants for that script. And then we took all of that input – I'm sorry. So the teams from the community ran the studies and did most of the work really, and the staff did the support, the logistical support.

And then what we did is we took all of that and on October of last year and we synthesized all of that into the Integrated Issues Report. And the report really looked at putting together an analysis – this is the Integrated Issues Report – putting together an analysis of the unique issues to each script, the common issues between the scripts, and it tried to highlight based on the information presented in the reports what further work needs to happen.

And with that I'll pass it over to my colleague Karen to talk about the specifics of the report.

Karen Lentz: Thank you very much. So what you see in front of you here is an outline of what's contained in the report. Most of the issues are discussed in sections four, five and six. There's also some project background and some descriptions of some types of Variant cases that were identified in the case study reports. And then also at the end there's a discussion of some additional work, which we'll talk about.

In writing a report about Variant issues it seemed useful to define what we mean by "variant." But when we, as we worked with the case study teams and looked at the discussions that had occurred on this subject previously, it became clear that there really wasn't a commonly agreed definition for the term "variant." It was used to describe a number of different concepts. And so while the report does continue to use the term "variant" in a loose sense, it tries to be specific where that's mentioned with an example of the type of case that's intended.

And it's recommended that when these issues are discussed, when "variant" is used, it's used with more specific information as to what's being referred to. And just to clarify on the scope of the report, the issues that are identified in the report concern IDN Variants at the top level. So we're talking about issues concerning IDN Variant TLDs. And then while there are variant issues at the second level that registries deal with, and a few of those are recommended in the report, really the intention is to focus on the issues at the top level.

The diagram here on this screen is a rough classification of the variant cases that were identified by the case study teams. It's not perfect, but we tried to take all the cases that were mentioned in the reports and classify them into some sets that would help to inform the discussion. So we'll go through some of these types.

The basic distinction that we arrived at was a distinction between whole string variants and code point variants. The code point variant uses the Unicode code points and refers to some sort of relationship or linkage between code point A and code point B or between a code point and a sequence or between a set of code points.

With whole string variants the relationship is between a whole string, so you'd be looking at whole string B and whole string A and identifying some sort of linkage or relationship between those strings; often in the context of those strings meaning to a particular language community.

So a number of the issues that were identified when thinking about how variant TLDs could be relevant to the root zone; there are two elements that would need to be established. You have rules for how you can establish what are variant labels; how do you tell if you have one or not. So the two elements that were discussed were a code point repertoire for the root zone and then a label generation rules, which I'll talk about in the following slides.

Establishing the code point repertoire for the zone, which in this case is the root zone, would include establishing the universe of which code points are permissible and which are to be excluded. Some of the issues with establishing

the code point repertoire include how do you decide which code points are going to be selected for inclusion, what sort of expertise is needed to make those decisions, and then how changes are handled if a code point is introduced or becomes valid or invalid; how does that repertoire handle changes.

And then the second piece that's needed for establishing variant labels is what we've called the label generation rules. This would include rules for how you tell which code points are to be considered variant of one and other, variant code points. And it would also provide rules for what status could be applicable to labels containing those code points. And I'll talk about the statuses shortly.

Issues associated with creating the label generation rules – there are a number of ways or processes that could be followed to arrive at those rules, which is a big consideration. Issues having to do with what expertise is required here, which label states are going to permissible, what flexibility should there be with use of code points with different script properties.

So assuming that you can establish variant labels, the next set of issues that we considered in the report was the treatment of variant labels. We identified six possible states that labels could have with regard to a zone. And obviously the user experience and the impact, the operational impact on parties using them is dependent on which sorts of states are used.

A theme running throughout the report was the user experience considerations and we talked about, or considered in the report, many different types of users. It was raised with us in some of the discussions that users have different capabilities with regards to a particular script from very familiar and uses the script all the time to someone who has no familiarity with that particular script.

So you can consider the range of different users with that in mind, as well as the different user roles. So we considered someone like a systems administrator or a network operator would have different needs or a different experience in working with variant TLDs than someone for example doing law enforcement to security work or an end user.

So looking at the kind of big picture, there's a little bit of a tension in the report that's reflected in the report between the great interest in creating functionalities and expanding the use of IDNs at the top level and the really difficult issues that are raised by what's trying to be achieved. The risks, if you think about security and stability, which is another thing that's discussed throughout in all of the issues in the report, the risks in that of the end user experience is significant if this is not carefully done. And then the costs are significant as well in creating a really sound approach to this. So it's important to weight the risks and the costs and take those into account to inform the decisions on future works that will be done in this area.

So I'll turn it over to Francisco Arias who will take you through the potential next steps.

Francisco Arias: Thank you Karen. So Naela and Karen explained the first two phases of the report that were focused on the issues related to IDN Variant TLDs – thank you. Is that better? I was saying the first to phases are related to identifying issues, now I am going to talk about what comes next. Defined in the final issues report that there are several potential projects that would need to be taken in order to be able to have solutions for the IDN Variant TLD question.

Currently the plan for these eight projects is in the public comment; the public comment period ends on the 18th March, which is next Sunday. Because of the new rules of the public comments that started this year, we will have a reply period until 8 April. That provided there is at least one comment received during the comment period.

Now I'm going to explain a little bit about the different projects that we have identified in the issues report. The first one is the label generation rules tool. This is basically the IDN Variant tables enhanced. The idea here is to have a uniform format that will support the different things that we saw raised using to identify the variants. For example, when we were talking with the Arabic case study, the Saudi registry was explaining that they have variants depending on

the position; that's a fissure of the Arabic script, which depending on the position of the code point you have a different shape, a different form of the code point.

So the current IDN tables that are out there do not have support for that. They developed their own solution and the idea is, as I was saying, to have a unified solution that can support those kinds of things. So this project is not really dependent on having variants in the root. This project is about having these tables that can be used by any registry in any level of DNS. And it's mostly about developing the specification. We are planning have it and are overseeing the idea. Kim is the author of this draft that is already out there and some of you might have already seen it.

We are here leveraging on the fact that some of us already go to IETF meetings and other meetings like the regional registry meetings where you will usually go there to talk with the registries and get input. And you can see here the estimate of the costs for the remaining of this fiscal year and the next fiscal year.

In this report we identified a series of feasibility studies that are needed to be done in order to determine which types of variants or states of variants are needed. One of those is project number three which is examining the feasibility of whole string variants. As Karen explained it before, this is one type or variant and in this case was requested by one of the case studies, the Greek case study. In that case they were asking to have what we call a dialectal variant, meaning if you have two different dialects of the same language they wanted to have two strings that represent two words in each of the dialects that by that language community was considered to have the same meaning. So this will be one type of whole string variant that they would like to have.

So in this project we are trying to define where it is feasible to have these types of variants in the roots and in the report we are defining several issues with having these types of variants. For example, you will need to identify the dictionary or dictionaries that you are going to need. And imagine for example the case of languages that are used in many regions or many countries around

the world. It may be really difficult to be able to decide which dictionary to use. And also the fact that you would probably need to do it for every language in the world is not something that seems so easy to do in a practical manner.

Other feasibility studies that we are considering is project number five which is the possibility of mirroring. Mirroring is probably the state that is most appealing to many people that are interested in variants. This is the idea of having two name spaces to map to one and other. So for example, you will have a stream in Simplified Chinese and a stream in Traditional Chinese and you want those two name spaces to map to one another so if you for example access a web site in the Simplified Chinese domain name, or Traditional Chinese domain name you will get to the same web page. That would be one way to see it.

The challenge here is that even if we were able to solve the problem in the DNS, for which there are already proposed solutions like using [D name] for example, which by the way we don't yet have that in the root so maybe that's something that needs to happen before we add that in the root. There is also another option which will be the [parlor] provision in the DNS, which is you have the two zones delegated but the registry makes sure that the content of those two zones is kept the same.

But the issue goes beyond that because you need to have mirroring work in their applications and this basically means all the applications out there. So you need the mirroring to work in the web for example in email, FTP, any other applications you can think of. And it's not really that simple. For example, in web, you need to configure the web server to know they names by which it's known in order to be able to respond to that. And something similarly needs to happen in email.

There are certainly some protocols that do not require that, but at least the two most common – web and email – need to be configured in order to respond, to know the names that they are known by. And you can see there the cost that we are considering for this study. This is basically we are thinking will be

conducted by two person, two experts a DNS expert and I can't remember the other expert that we considered. It's basically to have this study and we will be publishing this for public comment and have the input even in the plan phase and then in the study itself.

There is a third feasibility study that we are considering to have which is examining the user experience implications of active variant TLDs. By active variant TLDs we mean basically two states – the mirrored state which I already mentioned before and the other is the common delegated state. So this delegated state basically means to have two strings, again for example, the Simplified and Traditional Chinese versions of strings, delegated to the same entity but without the requirement to have mirroring in place.

So in this case we already know this will work from a technical point of view because basically that's what we have now. We delegate TLDs and there is no requirement to have mirroring. The question here is what are the implications in the user experience. What will the users be confronted with when they see these two strings that they may thing will give you the same result, but they may not give you the same result?

So this project we are considering to have a model similar to what we used in the issues report. We will have a series of volunteers with the community complimented with consultants. This is to have input from the volunteers. So hopefully at the end we will have, at the finish, what are the implications of having these two types of TLDs.

So once we are past these feasibility studies that I mentioned before, we will then, we're envisioning to have a Board decision on what types of variants and what states of variants should be implemented in the root. Once that decision is done, and we are planning for that to happen by the end of the year, then we will start on the work of defining the process to determine the variants; what we call the label generation rule set process for the root zone.

This process we are also thinking of building on the same model that we used for the issues report where we'll have a group of volunteers from the community

also complimented with experts we will contract to help us develop this process. And I should mention that we are considering this to be two projects basically. One is the development of this process and the second part is the actual filling out of the table let's say. That process, that second process which is identified here as project 2.2; at this point it's too far in the future. We don't know what the process will be, so we don't have an estimate of the resources that will be needed in order to have this done.

In any case the current plan is that this will happen in the Fiscal Year 2014. So we still have time to develop that. Similarly to the project two that I mentioned, we have project four, which is related to visual similarity. As Karen mentioned before, there was no agreement during the issues phase on what a variant is. So what we have basically is many types of variants and visual similarity was one of the variants that people identified during the process.

There was also observations of how this visual similarity process could be enhanced. For example, to use these IDN tables, which you will have identify the characters that are considered to be visually similar. So that you will know from start which are considered visually similar, so then hopefully you will have a better outcome on the visual similarity process. Again this is divided in two different projects.

The first one is to develop the process itself and the second part, which is to implement it. The implementation part is also too far in the future. We are unable at this point to determine what resources will be needed. And I should mention that in development of the process we are also thinking to use the same model as the issues phase; having volunteers complimented by experts.

And finally we have a couple of projects that are also in the implementation phase. So projects seven and eight we don't have yet a budget; we don't have an estimate on how many resources will be needed to carry on these projects. Project seven is basically to update the gTLD and ccTLD programs with any changes that are needed in order to support the label generation rule set and the visual similarity process.

Project eight finally is about to update the ICANN and IANA operations and processes that are already there. For example in ICANN I work in the area of the gTLD registry liaison; that area deals with gTLDs, they do the basics. So for example we will need to update the, have a CRM system that supports the concept of having a set of gTLDs which will be considered variants and treat them as thing, as a set, not to separate the strings.

Same thing will have to be done in IANA and maybe there is something to be done in their root zone management process, how the zone file is generated in root, etc.

This is a visualization of how the different phases, how we'll see the different phases. As I mentioned, we have the feasibility studies that we are planning to happen in this calendar year. And then we will have a decision point at the end of 2012 in which we are proposing that the Board will decide what types and states of variants would be implemented in the root. And then we will have the phase in which the processes are developed, pretty much during 2013. And then again another decision point to adopt these processes and then the implementation phase.

This is pretty much the same information just with a different projects listed in each of the phases. And with this I…

Dennis Jennings: Thank you very much indeed team. So, you've had a quick run through the work that's done, a quick run through the projects, and now the timeline is in front of you. Now that's what we have to say; now we're really interested in what you have to say. So the rest of this session, which can be as long as the time we have available, is in response to queries or discussion points that any of you would like to raise. We have a microphone in the middle there. Do we have a roving mic as well or are we expecting everybody to come up to the mic? I think everybody coming up to the mic.

So I'll continue talking like this until someone gets up and asks a question.

Chuck Gomes: Well I didn't want to listen to Dennis too long, did you? Chuck Gomes from VeriSign. First of all let me compliment all of those who prepared the consolidated report, plus all of those that made these huge contributions to make that happen. You're right Dennis, it wasn't an easy read, but I did make it through it – at least the main parts in the front. And I think it's in section five or six, somewhere in there. You did a really good job of talking about all the different types of implications and I appreciate that, it was really well done.

But one of them had to do with confusing similarity review in the new gTLD process. And obviously that's just about to happen. So I'm guessing that not too much of what's in your report can be taken into consideration in this first round, and that's understandable. But do you see any; I guess do any of you see any problems with that that could occur because of that fact. Now I know that IDN variants aren't going to be allowed in the first round, that's all really clear. But do any of you anticipate any issues that might result, because a lot of the work is going to happen after the evaluation of the applications in the first round because there's a huge task ahead in that regard. Just curious if you have any comments in that regard. Thanks.

Dennis Jennings: Thank you Chuck. You're right, this work on enhancing visual similarity in the IDN space is future work and it's not going to be available for the first round of new gTLD program. And it's beyond my competence or the scope of the project to comment on anything about the current processes in the new gTLD program, unless any of the team wants to comment, I don't think that's rally for us to. Any of the team wants to leap in? No? So we know that there is obviously an issue, but this project is not aimed at addressing that.

I should say there's more than similarity at issue here, it's confusability. I had a long discussion during the week with Lyman Chapin and he has some questions, I don't think Lyman is here, but he has some questions about how useful the visual similarity work will be to address what is really the underlying

issue of confusability. But that's a discussion for a later time. We also have people by the way remotely, so we will be taking questions remotely for anybody out there who wants to raise a question. But we have another question or a comment, please.

Male: My name is (inaudible); I'm talking about the program that we support in [Core], who I work for, which is .quebec. Quebec, as most people here know, is written the French language, which is the language of the province with an accent on the E. And it is written without the accent on the E if you cannot put it, or for instance in English. And it shows a very frequent problem that we have. Not only are variants sometimes needed for writing things correctly, they're also sometimes needed for writing things incorrectly because the incorrect way is the only one that is available.

I mean Spanish speakers know that. French speakers know that. When they type something they very often do not have the time, or the keyboard, whatever it takes, to type the correct decorated string. Now, in the case of .quebec for instance, we face a situation where applying for .quebec the wrong way, it's the wrong string, is going to be the correct string because that's the only one available for .quebec once it is actually running. And by not having the IDN string available at the same time, essentially we're teaching the population around the world, everywhere, we're teaching people "Don't you ever use IDN because it will not work."

Now for that to be taken back is going to take many, many years; it may even be impossible. If you look in Japan, which is one of the countries that has had the problem of not being able to write its own language in many, many computer applications for many years, people have become so convinced that it is impossible to write the native language that now there's even no demand.

So if you look at the IDN strings proposed for Japan you would be surprised, there are none, or almost. Because the population is convinced that it doesn't work; it's not possible to get it back. So there's a sense of urgency. We have to

do something, or simply after this round, in between to see if it can during the process over the rest of 2012 and early 2013, address these cases where the IDN could still be added to the string that is actually current in the platform.

Dennis Jennings: Thank you. There are many interesting thoughts behind that. First thing I'd like to say and then I'll get some of the team to comment, is of course domain names aren't words and domain names aren't part of a language, although a user perceives them as words. They are pneumonics and it's amazing what you can achieve with 37 characters, which is the ASCII set. And yes you are right that if you only have the 37 character sets, the 37 characters in the LDH ASCII set, people get used to that and that's part of the user experience.

And you're right that that's the only way that things can be done. In an ideal world we would have done this work a long time ago so that wouldn't arise. But the reality is we are where we are, and I know that's a terribly silly thing to say, but we are where we are and this leaves us some of the user experience consequences that arise. Now in the specific, who on the team wants to remind us exactly what the rules are about stating that you have a variant in the new gTLD program? Karen.

Karen Lentz: Thank you. What it says in the applicant guidebook is that variant strings, if the applicant believes that they have some can be declared in the application, which means that they're put on a list. They're not delegated or given to the applicant; there's no right conveyed by having declared the variant in that application. But it's essentially set aside until such time as a variant management mechanism is worked out for the top level.

Dennis Jennings: Thank you Karen. Have we any questions remotely Naela?

Naela Sarras: Not yet, thank you Dennis.

Dennis Jennings: Okay, you flag to me if there are. Next question please.

Wei Wang: My name is Wei Wang from CNNIC and I am now also the co-secretary of CDNC. And CDNC is that Chinese Domain Name Consortium that is in charge of the Chinese regulation about the registration and applications. And the members include CNNIC, TWNIC, SKRZ, which is .sk and SGNIC, [CONAC] and Afilias PR and Core etc. And CDNC appreciate that the ICANN set up their variant project and are trying to find a universal solution to the…

Dennis Jennings: I'm sorry to interrupt you. I'm having a little difficulty hearing you. Could you speak a little closer to the microphone?

Wei Wang: Yeah the CDNC appreciates that ICANN set up the variant project and they're trying to find a universal solution to the variant TLDs. But according to the [RIPE NCC] and the experience about the Chinese ccTLD and we'd like to say the implementation of Chinese variant TLD be moved forward as soon as possible.

Well actually yeah I also joined the Chinese Variant Report last year in 2011 and the definition of the Chinese variant has been clarified in that report and it was also established 10 years ago that there was no overlapping issues regarding to the [Han] character for Chinese, Japanese and the Koreans. So that means because of the Chinese variants have the same pronunciation and the same meaning, the Chinese speaking users, they regard the Chinese variants as the same and regard them as should be interchangeable.

And I think that's why the CDNC is involved in the publication of RFC 3743 and 4713 to try to – and the CDNC members have been providing the second level variant domain registration service over 10 years ago based on these RFC's and it was started providing the variant domain name registration under .china and .taiwan and .hongkong since 2009. And to date, no disputes are received concerning the Chinese variant confusion.

So because [of the renaissance] and the Chinese ccTLD operational experience, we think that the Chinese users priority expectation, or the first [the IDO] variant to labels should belong to the same registrant. And second the Simplified Chinese form and the Traditional Chinese form of the applied for [IDO] should be resolvable simultaneously or not resolvable at all. And I'd like to take .china, we call it [.xongou], both Simplified and Traditional for example, and currently we have about over 320,000 domain name registrations under .china. And among which over 77% have variant forms and mostly 20% of the DNS carriers for .china are for Traditional Chinese forms in China mainland.

So I go to so many data and issues that we have clarified in the Chinese Variant Report just to express that. We think based on all these experiences and the facts that we want to see the implementation of Chinese Variant TLD being move forward as soon as possible. Thank you.

Dennis Jennings: Thank you very much Wei and we fully understand that. Rather than respond to every question I'm going to take a few questions because I'd like it to be more of a dialogue between the members of the community rather than question, answer, question, answer. Because we're really here to learn what you have to tell us and to integrate that into possible a revised project plan of set of projects. But I will not forget your question and will come back to it. Next question please.

Suzanne Woolf: Suzanne Woolf and I think I'm going to help you out a little bit because I have more of a comment than a question. First of all, I wear a number of hats around here, and first of all as a member of the Board, who advocated very strongly for this project to happen, I want to thank everybody for all their hard work and I'm really pleased with the report and the results and that it's in front of the community.

When I first looked at this set of issues, one of the things that became clear was that it was being, there was an assumption within the community that there was a set of technical problems here and that all that really had to happen was that the technical people had to go solve the technical problem. And frankly I'm extremely pleased, it was basically step zero to question that, as far as how we think of the set of issues.

So I'm very happy with that, but another of the hats that I wear is as a technical participant in the IETF, among those technical people. And shifting to that for just a second, I have an observation about the feasibility studies, particularly regarding the technical feasibility of mirroring. We looked in some depth at that within the IETF, within the people that are best equipped to think about that, at least with respect specifically to DNS. And reached a number of preliminary conclusions about why it might be not only extremely difficult, but to a great extent many of the things that people want, and want for good reasons, are not just difficult but are impossible.

Because human beings aren't like computers, human beings need context. And the things that people are thinking about when they talk about a variant and meaning and words and all of those things really are not, can't be creating in the net; it's inside people's heads. Which is kind of a longwinded way of saying I have no objection to doing the feasibility studies, but people do need to be prepared for the possibility that the outcomes of the feasibility studies is going to be that at least some of what they want is not in fact feasible.

Dennis Jennings: Thank you very much indeed. And obviously we would very much like to involve the technical community so that we're not doing feasibility studies on things that people already know are infeasible. So that me shorten some of those studies, but thank you for that. Next question?

Thomas Narten: Yes, Thomas Narten here and let me just back up to the previous speaker who unfortunately has left the room. But on the example of like .quebec with two different spellings with the E for the accent and without. My sense is that that's not really considered a variant. And I would actually welcome a little feedback from the team here whether they agree here or disagree because one of the dangers of this space here is when we broaden the scope of what people are trying to do kind of well beyond what we even think of as being variant.

Dennis Jennings: Thank you. Francisco, do you want to respond to that?

Francisco Arias: Just quickly something. In the Latin case study, which was the first place of the issues case, said in their report that they recommended no variants for the Latin script, just pretty much in sense of what you are saying and they were saying it's better not to complicate things in the Latin space because we already have a [suitable] Latin script, which is ASCII, that is being used and it's probably not a good idea to add complication in that space, and I'm sure that's what you are referring to.

Thomas Narten: So I guess my answer is not – and I'll be a little bit blunt here perhaps – is that sounds to me a little like a wishy washy answer where you're saying "Well it might complicate it, let's not do it," this is, I think, a little more clear cut than that. Is the case of .quebec with E either accented or not considered a variant, because my impression is that in some context it might be and in other context

it's not. and with a variant, my impression of at least a single character variant is that they're always kind of equivalent. It's not like for some words yes and some words no, right?

Francisco Arias: Okay, let me see if I understand what you're saying. Yes, an E with an accent would be a variant in some cases and in others not. I could say I'm a Spanish speaker, I would not think you could consider the E with an accent and the E without accent as variants, but maybe in other languages it would be. I think that was one of the reasons why the Latin case study was saying you better not have variants in Latin because there are so many languages involved here that you cannot come up with a single definition for variants for the Latin script.

But in any case, in the report in the stage where we are with variant TLDs, we don't have any tables defining what are the variants for each of the code points. These are far in the future as you can see here. They are in the last phase, the implementation phase. That's where we will have a definition and that's depending on the process whether an E with an accent and an E without an accent could be considered a variant. But what it seems now, I'd say not very likely.

Dennis Jennings: Andrew can I invite you to comment?

Andrew Sullivan: Thank you. I just want to expand on this a little bit because one of the problems with this term "variant," I think you are pointing to it, it's that it's a little bit plastic. The point of, one of the points that we tried to make in this report is that because the root zone is a shared linguistic context for everybody in the entire world, the rules have to be consistent whenever you have a candidate variant relationship.

So the candidate variant relationship in this case would be an E without an accent and an E with an accentue. The problem in this case is that a lot of people who are using E's don't actually use the accentue's; that isn't the kind of variant relationship that they have. So what you would need in order to support variants in Latin in the root zone, is a set of rules that says "Every time you have an E here are all of the other possible combinations that you could have there," and that is the variant relationship that you would have.

Some of those relationships are in fact inimical to other uses of other linguistic communities and that's actually what the problem is. So that's the reason that the Latin study says "It would be a mistake to implement variants for Latin in the root zone." This doesn't mean that in a zone that is directed primarily to French speakers it would be wrong to have a relationship between an E without an accent on it and the accentue and the accent[frav] in some other cases.

It might be legitimate to do that in a different zone, but in the root zone, everybody has to participate and everybody's problem has to be addressed. And sometimes the answer to that is "Because we cannot satisfactorily do this in a way that is technically feasible, it's logically possible but it's technically feasible, therefore we shouldn't do it at all."

This is the same reason why in fact, despite my sympathy, and I'm speaking personally here, despite my sympathy with the concerns of in particular the HAN using community, with the pace of change here, the problem is that we need at least a framework for a set of rules by which we would understand how to process those variants. And that framework needs to work for other linguistic communities too. So the fact that HAN users are ready to go isn't actually enough to do this in the root zone in any kind of systematic way. And it's that systematic way that is the problem.

From my view, and again I'm speaking personally, we could do this on a case by case basis, but that just turns it into an endless political struggle right, every case has to be looked at. So what we really need is some kind of mechanism, and it might be that the mechanism could be developed more quickly for HAN

characters, but we need a framework that actually would accommodate these other cases as well before we're ready to do that. And I think that is one of the projects that is on this list. If we think that that framework development is something that's really urgent, maybe it needs to be moved elsewhere in the project plan, I don't know. I don't have an opinion about that. But if people have an opinion it would be nice to hear it.

Dennis Jennings: Thank you Andrew. I think you've articulated very well some of the things that we've got to struggle with and some of the concerns that the community have when seeing this timeline. I understand we have a remote question, or a question from a remote participant. Naela?

Naela Sarras: Thank you Dennis. We have a question from David Cone and it says "Several existing registries intend on applying for transliterations of their existing gTLDs. Do the findings discussed here have any effect over these plans assuming, for the lace of a better word, that only one other/additional version and wanted by the existing registry."

Dennis Jennings: Yeah, who wants to take that question? Francisco, Andrew, Kim – Francisco has lent forward. Go ahead.

Francisco Arias: So, translation or transliteration of streams are other types of whole string variants, since you cannot do any of those code point by code point you need to consider the whole string as a whole. And you also probably need to consider some language background in order to do the translation or transliteration. So this could be considered as a whole string variant possibly too, sorry.

Dennis Jennings: Thank you. I have to say, speaking personally, I am skeptical that we'll ever find a deterministic way of dealing with whole string variants. But we need to look at it. Andrew, did you want to add anything to that?

Andrew Sullivan: No.

Dennis Jennings: Okay, next question please.

Edmon Chung: Thank you Dennis. This is Edmon here. I guess as a long term participant in IDN policies and technologies I first of all wanted to congratulate the team and send, applaud the team on a very good set of case studies and a very nice framework that was put forward in terms of the integrated report. I think we, going back to actually – oh before that I want to say that it feels a lot, this morning in some of the discussions in the last month or so, feels a lot like déjà vu. I saw a lot of these things that have been talked about for a decade.

And I think I'm glad that we're here in this forum and be able to address it and go into an implementation mode in terms of at least, I think the direction we're headed is to implement. And what Andrew just said I think it very important, but I would like to think about the whole project as a practical implementation more so than a theoretical approach.

And the reason why I say that is because of the decade long discussion there is a critical community consensus that I think was the basis of the success of the VIP team so far. And that is that we split it out into different case study teams and different script and languages. There are certain things that are unique to certain groups of languages or groups of scripts. And if we try to lump it all together, I think we're looking at a project that will never end. And that is not practical I think.

And if we have a mind towards implementation, then we should come back to practicality and look at what, as I forgot who's mentioned, cost, benefit and risks. It's a balance. It's always a balance. And I think how we balance it, the community already, in terms of the discussion of a decade, more than a decade long discussion already has a community consensus. And that is that we split it out in to certain languages and scripts.

And I think right now, what was found in the integrated report, I think is very enlightening in that the framework as Andrew said, sort of a universal framework, is already there. The project plan, the eight projects – that is the framework. That is the framework for each language or script to move forward on. And if we, basically the suggestion and I think I urge the team and I wonder if anyone else from the community will add to that, is that we can use what has successfully been created from these case studies into an integrated report, which is a universal framework, and proceed based on scripts and languages, or groups of scripts and languages.

And I think that will lead us into practical implementation. And from what I hear just know, from Wei Wang from CDNC, 20%, 20% of the queries of the IDN variant TLD for .china is for the variant TLD. That is a significant number. That shows significant need from the community. And we cannot, and we should not deny that. I think that is really what the team needs to focus on, and especially going forward with the project plan.

And I think splitting it up in contained – you have to balance risks. I understand that there are risks – cross scripts, there are different types of issues, but we have to balance risk in terms of technical risk as well as Andrew mentioned, political risk, as well as users. Ultimately I think this is done for end users. And 20% of the users are using variant TLDs today. So that I think is really a very important aspect that the team should look at. Thank you.

Dennis Jennings: Thank you and I agree fully with you. But, and I'll hand over to Francisco and Andrew, first of all a top level question. Yes there's been a lot of discussion

about this but I don't believe this – I believe that this is the first time that a timeline for addressing and resolving these issues has been put up, so I think there's been huge progress.

I know there's frustration that this looks like a long timeline and I know there's a great need. But there are some things that we need to comment on, and I repeat I think, about the root zone being a shared unique resource. And in fact that dividing things by script doesn't quite work because what's required is more than one script as defined in IDNA 2000 and in Unicode. Francisco, do you want to pick this up and explain why we think it's a little more complicated than that, and then maybe Andrew?

Francisco Arias: Sure. Thank you Edmon. I just want to say that we certainly understand what you're saying, the need for certain communities to have variant TLDs. But we are also taking into account that this is a shared root and we need to see the big picture here. And I must also say that we are not trying to not use the model of the case studies. But there are some things that are common to all of the scripts, or all of the case studies. For example, many of the case studies identify they need to have mirroring. You need to take a look at mirroring from a specific strict point of view; that's the same need for everyone.

So you can do the feasibility study about mirroring without taking into account the scripts, because it's the same. Something like that is true for [break sys] about variants with and without mirroring the user experience. That's also something that was requested by, that is probably required by different scripts. So you can take a look at that independent of the script.

The part that I think that is dependent on the script is the implementation part, which is the third phase here. For that certainly, or more likely we will need to have a specific theme or specific solutions for each language or script community. Filling out the variant table is certainly that probably the Chinese community would like to have done by themselves, the Arabic community and so on. So we are thinking that that should be done in the phase where the issues

are specific to a language or a script community, but there are other parts in which the solutions need to be think for the big picture or where there are shared concerns.

Dennis Jennings: Yes please. Let's have a bit more of a discussion than…

Edmon Chung: Sure. Well I guess it goes back to again risk, cost and benefits and how you balance that. I understand those issues. And the specific issues as you said, and in fact I go back to both Thomas and Suzanne there that some of these issues perhaps, like the whole string variant or like the mirroring could be dealt with very quickly. And also if you look at the case study reports, each of them have varying degrees of their request for that. So by splitting up into groups of scripts or groups of scripts/languages and I'm not talking about anything in particular. If you do that than some of these go away immediately right.

For example Chinese, mirroring goes away because we already have the experience right now. Whole string variant goes away. We quickly cut down on the path towards implementation. And that is exactly why the community consensus was to separate it out. Yes I understand that there are cross cutting issues, but once you split it up then some of these issues go away or make it easier for that group of scripts that have that implication to move forward. And that I think is the reason why that's more practical.

Dennis Jennings: Thanks Edmon. Andrew, do you want to pick up this thought about? Oh, Naela?

Naela Sarras: Thank you Dennis. I think I would like to make just one small comment for Edmon and everyone is that part of putting this plan out for public comment is

that we're really here looking for feedback on – one is how can we work together to maybe improve the timeline and two is how can we prioritize these projects too. Because we're here with the community, we want the same things, so we're asking for help with prioritizing. And if there ways that we can talk about shortening that timeline; we certainly want that dialogue to happen.

Dennis Jennings: Thanks Naela. Andrew, do you want to comment about taking sort of a subset of this and moving it more rapidly through the process as a sort of parallel set of subprojects.

Andrew Sullivan: Yes, thank you. I just want to draw people's attention to section 4.2 of the integrated report, because it actually talks about how you make decisions about the label generation rules. And this was precisely my point earlier, that that's a framework that everybody is going to have to live with. The problem that we have here is that we've got a range of possible ways to do this. One of them is to just take the Unicode code properties and hang everything off that. We didn't actually do that in the project, because the case study projects didn't restrict themselves always to the same script, and they didn't always cover the entire script.

Frequently they didn't have enough expertise to cover the entire script and in at least one case of course we had a group that was, strictly speaking, talking about Chinese, even though that script is shared with other languages. Now there was some discussion of other languages, but the fact of the matter is it focused mostly on Chinese.

So, if we're not going to hang this all on the Unicode code properties, which is one extreme, we have a number of other decisions to make about how we're going to draw the boundaries around the code points that we're going to be working with. We need to make that decision and we need to make that decision irrespective of the language communities urgent, and I admit, I

completely agree with you that it's an urgent and overdue need and there's been a lot of discussion for a long period of time that has not been systematic enough to get us here.

It's too bad that we're at this point of view, but this is the way the world is. And if we're going to just make ad hoc decisions then we are going to cause problems for people who simply aren't in this room. There were only six cast studies there. That is a tiny fraction of the users of the various languages in the world and we haven't begun to cover all of those cases.

Now, some cases are a little bit easier than others, but if the answer is "We're not going to follow the Unicode code properties," then we need to make some other kind of decision and we need to make that decision in a way that allows us to have a mechanism in the future, so that when somebody else comes along and says "My community isn't represented here," we can say "but we have a mechanism for that." The alternative is that the root zone, the repertoire of code points that we allow in the root zone become subject to sort of annual political fights about who's going to get in here next. And that is irresponsible management of the root zone, in my opinion. Again, I don't speak for ICANN so I can say these kinds of things.

That would be really, really a serious mistake in my opinion. We should not make that kind of decision. We should have a mechanism that we believe that we can work with in the future. That doesn't mean that everybody's problem is always going to be solved in the future. There are always going to be compromises here because this is a shared system. And I'm not advocating that sort of everybody has to wait until the last person on the internet is ready to go; that would be an insane answer.

But we need at least the meta-level of procedure by which we can say "I know how I'm going to so this subset of this thing," and we need to figure that out. That thing has to happen early. If the answer is this project timeline doesn't get that thing moved early enough then I think that is valuable feedback. But the community needs to make a decision about what compromises they're willing to

put up with in the future when somebody comes and says "You did something that disadvantages me," because at some point we are going to be faced with that and we need to face it.

Edmon Chung: Just a very quick comment. And that is exactly why we have this particular integration work and the first phase of VIP is supposed to answer that question, at least in my mind, and I do believe we already have the framework and that's the result of this phase of the VIP. We might disagree on that but I do believe that we already that framework based on the results here.

Dennis Jennings: Thanks. Next question or comment.

Female: Hello, this is (inaudible) from [CONAC] China. So we're applying for [.jungwen and .xongie], and that's two Chinese script strings accommodating the Chinese Government Organization Communities and the Public Interest Organization Communities. I think the need for clarification to trace and traditionalize the Chinese and simplify Chinese is really crucial here. Because our potential registrants they treat equally the end users that is using traditionalized and the simplified scripts, so there should be no barriers when we are offering the services.

We hope that ICANN can see the active influence that might be exerted by delegating only Traditionalized Chinese or Simplified Chinese. So it is really frustrating that ICANN comes back to the technical issue time and time again. There is a consensus here for the case of .china and for the Chinese community, even for the Asian community here. An applicable technical solution has been there for a long time as you have just heard some of the dataset provided by the CDNC. And the best practice for Chinese Variant TLDs has already been developed and proved to be satisfying to the end users, to the Chinese communities. You have heard that repeatedly, I believe so.

So here once again we are asking ICANN to give us priorities. We are asking you to prove that the case by case solution is not something you keep talking about, but not something you try to do. And especially we would like you to give us a priority because we are really well prepared and we always have high expectations for this for such a long time. We've waited too long. And I hope you can give us some serious consideration about that. Thank you.

Dennis Jennings: Thank you very much indeed. I think there is a common theme coming here, and we fully understand it, and I think it would be very helpful if we had some way of having a more intensive dialogue about whether in this framework it is possible to look at the Chinese case as a separate, as an isolated case. But we certainly aren't going to do that on the fly here at this meeting. But we do, I mean we do really, really understand the need.

But the community also has to understand the responsibilities that we have here in not screwing up, and in particular – that's a technical term – and in particular not unfairly disadvantaging other communities because remember, there would be an element of first come, first served here. And those communities that use a script in their language may find that the way the code point variants are defined, it does not actually work for them and they have a set of label generation rules which are inappropriate for the language. Now of course there's always going to be an element of first come, first served and the judgment is how to get that reasonably right. Next question.

Thomas Narten: Thanks. Thomas Narten again, and let me go back to, I think it's – is it recommendation three or project three, the single, the whole word? I forgot the numbers are not in order. That's okay. Examine the feasibility of whole string variants – I guess let me express a bit of skepticism of this. And the question I have is, is this being proposed for completeness, or is this being proposed because there's a real need and there's a strong belief that this is valid form of what people consider to be variant?

And also is this work that is sort of somewhat independent of the other projects in the sense that there's not necessarily dependency on going through this and working it all out. I guess to go back to what Edmon is saying, I think the thing that needs to be clear to everyone is what are the critical dependencies for moving forward in general, and what are things that would be useful to do that would inform the debate but could also be done kind of in parallel in the side that are not critical.

And then you could ask the question like "maybe we should just wait and do those later; maybe we should deprioritize it and make sure that we focus our resources on the critical dependencies."

Dennis Jennings: Yes absolutely. Francisco, do you want to comment on the whole string question?

Francisco Arias: The reason why we have this part in here is there was one case study that insisted that this was a need for them. However in the report we are raising serious concerns about if it's really the help of having whole string variants in the group. So I could say I hear what you're saying. I certainly think it's a very difficult to case to make for a whole string variant in the root. But the report is an issues report, but what we did was report what the case studies told us. And one of those told s that they needed whole string variants; that's the reason why it's there.

Thomas Narten: So let me just reiterate the point that I make is that I think it would be useful to prioritize resources going forward where the critical dependencies are, with an eye of – if you look at the big picture you want to get to the point where we actually can start talking about seriously delegating certain languages and certain cases. What are the critical dependencies that have to be done before you can get to that point?

Dennis Jennings: Let me just pick up that and sort of put it in more layman's terms. If we pulled out this project and did it later would it make a difference? Right? Okay, that's a clear question that we will consider, look at and address. Naela?

Naela Sarras: I have a comment from online, may I make the comment? And this comment Francisco, is about the timeline, so maybe we can advance the slides. Okay thank you. So the comment comes from Joseph Yi – he says "For several projects studying the feasibility is not clear. I'm not asking to have it defined now, but defining criteria of what's feasible or not feasible helps not just timeline and priority, but helps for the big picture for impact."

So I think if I understand Joseph Yi's comment is we could probably maybe elaborate better on what we're proposing in studying the feasibility.

Dennis Jennings: Yeah I think we need to elaborate – although they are in the project plan and people should read it. We elaborate what the projects are and the rationale for them and we're looking for feedback on that. I think one of the things that arises from this is that – I don't want to put words in people's mouths but we certainly have had feedback that the closing of the comment period next Sunday is unreasonable and it needs to be extended.

If anybody wants to make that comment from the community and read it into the record that might be useful and we are certainly considering – Edmon, go on. I know you want to read that into the record, go on.

Edmon Chung: I think most of the community feels that there was a four week comment period you put out and it was right before the ICANN meetings and ICANN week and I think at least, from the discussion that I've had in the last couple of days with

the community, I think an extension of at least about 21 days would be useful for us to give you constructive feedback.

Dennis Jennings: Do people agree? I mean I think that's a very reasonable request. Does the community think that this really needs to be extended so that it can be properly considered by the community, commented, discussed and an appropriate project plan, a community project plan be brought forward – I think it's overwhelming, the case, so we will bring that back.

Alright, next comment? Any comment from the panel here that they want to make. Steve, any comment? Andrew? Kim, tell us how your project is going.

Kim Davies: Well, one of the tangible things we've actually started on, as shown on the timeline, is the label generation rule set tool. And the way that project has commenced is initial drafts of an approach to an XML based representation of IDN tables. How this will play into this project is obviously yet to be determined. But we suspect that whichever solution for variants reaches the finish line, so to speak, there will need to be some form of ability to take code point sets, variant sets and compute them into candidate strings.

We've also received independently a lot of feedback that the current approach ICANN takes to IDN tables will be more beneficial if it's in a consistent machine readable format. So recognizing the benefit of this we've started some initial work on this. There's already an IETF internet draft published as well as some sample code. And also a conversion of most of the existing IDN tables in existence to this proposed new format to stimulate discussion.

So that internet draft is out there right now. We certainly welcome feedback and participation. And also, I know I've spoken to a number of TLD registries that are here this week and had some very useful dialogue about how they use IDN tables, how they could benefit from this format and that input will feedback into future revisions. So I'm certainly very interested in feedback on that technical

specification, and anything we can do on that particular element to help support usage of IDN tables more generally, obviously for this project it's about the root zone, but I think the application of that specification is more universal than that.

So we're very willing to hear feedback on that.

Dennis Jennings: Thank you Kim. Comment, question?

Male: This is a question that's a new question, it's not referring back to what Kim was just saying, so I should really let anybody who wants to…

Dennis Jennings: Please go ahead. We're very close to time, so please go ahead.

Male: Yes. Okay, right well what we've heard this morning, I mean you just feel as if you're between a rock and a hard place because you've got the really solid technical arguments on the one side, you've got passionate arguments coming from, actually largely from the Chinese community, but it could also have been from one of the other really big communities like the Russian community or the Arabic community perhaps.

And so we are in a very, very difficult position here and what I am going to suggest is a British compromise and say that actually what we might want to do, we know the Chinese community is extremely articulate, they are obviously very frustrated, so this is a community we really need to prioritize. I think probably the Arabic and Russian communities are likewise, because these are communities with huge numbers of users, possibly the Devanagari community as well.

And we really do need to be looking at particularly solutions for those languages because I think that will really take a lot of the heat out of this. Now, fortunately the problems that these various scripts present are very different. So I won't go back into the Chinese issues because I think that's already been very well articulated by the Chinese community.

The Arabic issues came up briefly because we had this mention of positionality, so Arabic is a script where basically the same letter has about four different forms depending on where it comes up in the word. It's also got lots of possible problems of visual similarity because there aren't actually very many letters in Arabic, but depending where you put dots then you make the difference between a T and a B and this sort of thing, so that has a different set of problems.

And then Russian, yet again has very, very different issues. Probably major visual similarity issues with the Latin alphabet. For example there is now an application for [.muskva], I don't know whether that will be successful or not, that's the Russian Capitol Moscow. And of course it just so happens that the letters for [muskva] are exactly the same as [mukva]. So we are no longer dealing with a theoretical overlap problem, we've actually got a real overlap problem there.

So what I'm saying is that perhaps we could do worse than prioritizing those really big languages, see if we can make progress with them. It doesn't mean that we're not listening to the smaller languages and there may even be a case of talking to the Hebrew community and the Thai community particularly. But perhaps that could be some kind of a way of reducing the heat here.

Dennis Jennings: Thank you very much indeed for that helpful comment, which we'll take on board and see if we can – the message is coming fairly clearly and we're going to have to try and consider whether there is any way that we can accommodate to some of the pleas and the passionate pleas that have been made by some of the community.

Male: I mean, presuming this sort of approach were taken it would mean that we would be particularly interested in any problems that are not thrown out by those majorly different scripts. I mean there may well be some, so we're not saying there aren't, and those are the things we now want to hear if we would go ahead with that sort of approach.

Dennis Jennings: Thank you. Back to the team, Karen, and comment? Naela, Francisco, any comment? Right, we're one minute from time. I think I'd like to emphasize how much we appreciate the input from the community. I'm going to ask the team up here, on behalf of ICANN, to applaud you as representatives of the community; team. Thank you very much. This session is over.

[End of Transcript]