GISELLA GRUBER: Good morning, good afternoon, good evening. Welcome to the APRALO webinar on IDNs advancing the internationalized domain names agenda for a more inclusive Internet on Tuesday the 27th of July at 08:00 UTC.

As this is a webinar, we will not be doing a roll call, but all attendees will be noted on the agenda Wiki page.

On today's webinar, we have English and Chinese interpretation.

With this in mind, please state your names when speaking to allow the interpreters to identify you on the other language channel as well as for transcription purposes, and please do speak at a reasonable speed to allow for accurate interpretation.

With no further ado, I will hand the floor over to Satish Babu, APRALO chair. Thank you, Satish, and over to you.

SATISH BABU: Thank you, Gisella. Good morning, good afternoon, good evening wherever you are. I do notice several from outside our region. Welcome to everybody. As Gisella has mentioned, this is a webinar on IDNs, and the subject there is advancing the internationalized domain names agenda for a more inclusive Internet. And that's an important outcome that we are seeking, and IDNs form a very important step in that.

I'm Satish, I'm the chair of APRALO, and this webinar is organized by the APRALO policy forum and supported by the ALAC Capacity Building Working Group.

For us in Asia Pacific, IDNs and universal acceptance are very important topics given the vast diversity of languages and scripts in the region. And as the ICANN community works towards a new round of gTLDs, one of the areas that we need progress is on IDNs.

Now, the long work that the community has done in the SubPro PDP working group, actually, the IDNs have been flagged as one of the recommendations, number 25, and the newly launched EPDP will further take forward these issues that have been identified.

So given the fact that the new PDP is about to take off, we're at the cup of this kind of development, and the APRALO leadership team, with some input from Justine who initiated the whole thing, decided that we want to organize a webinar, inform the At-Large community and anybody else that's interested in the whole ICANN community beyond on where we are and what are we trying to do, what are the next steps and so on.

So today, we have a panel of three eminent speakers—of course, Justine [inaudible] organize it—who would be talking to us on these topics. On behalf of the APRALO community, I'd like to welcome Emily Taylor and Sarmad Hussain to this webinar. Both are very well-known figures to us.

Emily will be speaking to us about the basic definitions and general landscape of IDNs and also highlight some of the trends. She will provide us with some insights as to what are the impediments or barriers in the adoption of IDNs.

Sarmad will take us through the [inaudible] IDN program in ICANN and also some of the under the hood technical aspects and challenges in the

context of implementing IDNs. He will also talk to us about IDN variants and the kind of challenges they pose, and provide us some context on the EPDP on IDNs.

And finally, Justine will connect for us the work done within the SubPro PDP working group, connect it to the new EPDP on IDNs so we're able to see the connection between these two. Now, this is an interactive session, it is meant for our participants, and any kinds of questions or comments are most welcome.

We can type the questions and comments in the chat box. Gisella would be doing it if she has not already done on how to raise a question in the chat. We will take all the questions after the presenters have completed their presentation. Lianna and Amrita who are the vice chairs of APRALO will help us in curating these questions from chat and they will also be backup moderators in case I get dropped.

So with this, I'd like to once again thank all the speakers, and I'd like to request Emily to start with her presentation. Over to you, Emily.

EMILY TAYLOR: Satish, thank you very much for your introduction, and I'd like to also thank Justine for inviting me here today and for organizing this session. It's always a great privilege and a pleasure to speak with members of the Asia Pacific region about IDNs, because as you said in your introductory remarks, Satish, for this region with all of the scripts and many languages that exist, the rich array of languages that exist, you don't need to make the arguments that you perhaps would in other contexts of how necessary having a domain name system that speaks

everybody's language is. And not just for nice-to-haves but as an essential part of using the Internet to its fullest extent.

We've seen in recent years that apps and social networks and many applications have given very widespread language and script support, and yet the usage of domain names in different scripts and languages is really lagging behind.

This is an impediment to people's navigation, to finding resources in search. A very interesting piece of research a few years ago shows that when people have search results, that they really do use the domain names as an additional cue to understand whether or not they should click through the link and whether that really is what they're looking for. But of course in business or organizations, the use of e-mail and websites is still an essential component. Next slide, please.

Thank you. So at a basic level, what are internationalized domain names? Well, as many of you will know on this call, the traditional character set supported by domain names is very narrow. It was the A to Z in Latin script, zero to nine, and the hyphen. Together, these are referred to as the ASCII domain names, and that's what I'm going to do today.

And at a technical level, the IDNs transliterate Unicode labels into ASCII strings. So, what does that mean? I've shown on this slide examples both of second-level domains, so we've got paradigma.eu, you'll see that the meaning of the second level IDN is that it is a different script occurring against or inside at the second-level within an ASCII top-level domain.

And you can see that what the human is seeing in the top example is the paradigma.eu, so Greek script, with the accents and the diacritics and the different script. But what the computer understands is still ASCII. So that paradigma would be transliterated into xn--blah-blah. And good luck, if you're a human, trying to remember that Punycode string.

At the bottom of the screen, we have what are called top-level IDNs. That's a much more uniform user experience, particularly for domains that are not fundamentally based on the Latin script. So here we have an example in Cyrillic. And the user will see a fully Cyrillic domain name under the RF, which is the Russian federation ccTLD IDN. And then again, you have both components translated into Punycode underneath.

So IDNs have actually been around for a very long time, more than 20 years. We started to see them rolling out in .com and .net. And within this region, we believe that .jp was the first to offer IDNs at the second level back in 2001. But I'm very ready to stand corrected on that if anybody's got additional information.

And at the top level, the second example that we have here on the screen, those became available through primarily the ccTLD IDN fast track done by members of the ccNSO within ICANN. And that happened from 2009 onwards, we started to see the first ccTLDs roll out, and then of course with the gTLD program, there was about 100 new top-level IDNs. Next slide, please.

So I'm here today on behalf of the partners of the internationalized domain names world report, IDN World Report. It's a project that's been going on since about 2011 we started studying the rollout of IDNs, and

it’s been supported all the way through by EURid and also UNESCO, and a couple years ago, we welcomed the ccTLD RU as a partner to the IDN team. You can find our report at idnworldreport.eu and we’re just starting to roll out the figures for the 2021 world report on the site. Next slide, please.

I'm just going to talk to you about the key findings—we're in transition, so there'll be a bit of last year and a bit of this year—and talk about some recent studies that my research team at Oxford Information Labs have conducted on universal acceptance and some coming up. And then just a quick look at IDNs in the region. Thank you. Next slide, please.

So this is really looking at last year’s report. And on the left-hand side, we can see the percentage growth rate of all domains in dark blue versus IDNs which are in the sort of turquoise color. And what you see immediately is that the growth rate of IDNs is much more volatile than that of the regular domains. Lots of peaks and troughs. High growth about ten years ago and also negative growth in two of the last three years. And early results for the end of last year, 2020, show a plateauing. It’s a slight reduction but not a dramatic one.

Now, there are a couple of reasons for that. Firstly, they're newer than all of domains, but also, we’re going from a much lower base. So according to our report, we find between 8 and 9 million IDNs in the market, and that’s across ccTLDs and gTLDs. So of course, a small change will result in a larger percentage change than you would see from the base of about 300 million domains in the whole world. Next slide, please.

For this audience, you will of course know the difference between generic top-level domains and country code top-level domains. But on the left-hand side of this chart, what we see is the split between the two. And the immediate message is that the majority of IDNs are found in the country codes. And that in a way is not a surprising aspect, because the country codes are very much rooted in their geographical area. They support the language sand scripts of the people who use them and depend on them.

And some research that we've done over the last few years both on the language supported—we looked at the content of language, of web content associated with IDNs in ccTLDs, and we really found that we were looking at Europe, so the Dutch registry, and this was across the Board, actually, not just IDNs but their whole domain, you would see English and Dutch in that country code in the web content. And that really does seem to be the pattern. The country codes really strongly associated with the language, the culture and the scripts of their region.

At the same time, we've seen peaks and troughs in the gTLD space, and maybe we can hope that if there's a new round of new gTLDs, that there would be more adoption of IDNs and maybe we can keep that thought going through the discussion.

At the script level, on the right-hand side of this screen, we see that really four scripts are responsible for the majority of IDNs. Han script, which is associated of course with Chinese language and is a component in many Japanese words as well, Latin script, so that would be Latin script with accents and diacritics as you would see in languages like French, German, Spanish and so on. And also Cyrillic associated with

Russian, but also many of the CIS countries. And Hangul associated with Korean.

Now, in that other bit, you've got lots of others. but when you think that in other, that small percentage, you have major scripts like Arabic, Devanagari associated with millions of speakers. You can see that even in the IDN world, there is a real concentration towards not just the major languages because these are major languages, but they're not really are represented in the domain name system to the degree of their offline use. Next slide, please.

Last year, the year of this report, we also finally saw the launch in .eu in Greek. I think that is a video which is why it's blurry. But the .eu in Greek script was a very long journey that Sarmad and I have lived through in the last decade that finally, it was launched in November 2019. And that brings, for EURid which is one of the lead sponsors of the report that brings its suite of top-level domains into completion so that it supports all of the scripts and languages officially spoken in the European Union. Next slide, please.

One of the things that we find is that IDNs help to promote linguistic diversity in cyberspace. And what this map does is it looks at the location of IDNs and colors each one according to the script. And the result is more or less what you would expect to see if you were guessing, which is that in Europe which is very dependent on the Latin script, you see Latin script IDNs, similar in Latin America. In the Middle East and North African region, you see Arabic script. In China, you see Han. In Korea, you see Hangul. And in the CIS region, you see Cyrillic. And the

sort of mustard color represents mixed scripts. Thank you. Next slide, please.

This really develops a scene of IDNs and linguistic diversity. And I know that the labels are very small, but there are three bars here that we're seeing from top to bottom. The first bar shows the primary language of people in the world, and what you see immediately is that English is representing between 5 and 10% of people's primary language. It's of course a second language for many people.

Chinese is a large language, as is Spanish, Japanese, Portuguese and so on. But his light blue of other languages represents about half of the people in the world. And that really supports UNESCO's finding which is there are about 6000 languages spoken today in the world. They think that within 50 or 100 years, half of those will have died out.

Now, when you go to the second bar, we're looking at the primary language of web content. And we get this information from the W3 techs who run an analysis each year. Unfortunately, rather discouraging finding is that the primary language of web content is actually growing—or rather, English is growing as a proportion of primary language of web content. It is now 60% on the latest finding, but in previous years it's gone up from 50%. So the story of web content is very much dominance by one language.

What we do each year is we look at the language of web content both by crawling the gTLD IDNs and assessing the language ourselves, and also by making inferences and through talking to colleauges within the

ccTLD community and based on our research of ccTLD zones when that's been available.

And what we see is that English is still overrepresented with about 9% of content in IDNs, but it is much more in proportion with the offline language speakers.

Chinese and other major languages are also featured quite heavily. But unfortunately, again, we do see languages like Arabic still very poorly represented even within the ccTLD in the IDN world. Next slide, please.

So, what's going on? I hope the main bulk of our conversations today will be about universal acceptance, this term which is a little bit confusing. But really, it means, how are they to use? And unfortunately, throughout the period of our study, which hgo3es back to 2011, we have found that universal acceptance, the usage of IDNs, has been and remains very limited. And what little gains there are are quite slow, particularly when we compare the accelerating rate of technologies in other areas.

So, why is this? Next slide, please. I'm just going to take you, at a very high level, through a study that we did last year for EURid. And we were looking at e-mail address internationalization both in sending and receiving and replying to a selection of IDN e-mail addresses. These were all tested in European languages because it was a study for EURid. Next slide, please.

This is kind of a nerdy point but I can see that there's probably some interest in this group. We found the most interesting part of the study was the challenges we found in setting up our control environment,

which we thought was going to be the easy bit. We had hoped to use cPanel which is quite a popular hosting solution, and of course, this is relevant because for registrars who are selling IDNs, they will want to be able to support them as well.

But we found that cPanel actually didn't support IDNs adequately enough to set up the control environment, so we kind of made our own custom Postfix and Dovecot configuration which gave the full support for Unicode at IDN sending and receiving. So I think that was kind of an interesting aside for us about the general environment and the failures that you see in quite low-level applications. Thank you. Next slide, please.

We found that there's reasonable support in most of the e-mail clients that we checked for IDN e-mails, but there was also quite a lot of failures. And popular e-mail clients still have a long way to go to meet the mark for universal acceptance in e-mail. And when you consider the beginning of this session, I said that IDNs have actually been on the market for 20 years. So e-mail is probably one of the two things that you would immediately think of as a way of using IDNs. So that’s perhaps a surprising, slightly disappointing finding. Next slide, please.

This year, we’re very proud to have been selected by ICANN and Sarmad’s team in the Universal Acceptance Steering Group to do two further studies on universal acceptance. We’ll be looking at social networks and also web browsers. And those studies are ongoing. So if you're interested in that, do keep an eye out for those results, which I hope will be later this summer. Thank you. Next slide, please.

So just in closing, I'd like a quick review of the Asia Pacific IDNs. Next slide, please. The first thing to note is that the majority of IDN registrations are in this region. So in that way, the geography of IDN registrations is very different to that of traditional domain names which would tend to be seen in North America and Europe. Here we see the big centers of IDNs are in China, in Russia and other Asian countries. Thank you. Next slide, please.

This is quite a confusing graph, so apologies for that. A rough and ready view of the data that we have by country in the region of the IDNs. And you'll see immediately that there are gaps. So if you look at the blusih grayish one, that is IDNs in China and you see that there are gaps int eh data between 2016 and 2018 and that there are also very big peaks and troughs.

So the mauve-y color at the back is Vietnam which operated a free to register scheme for its IDNs for several years and then when it started to charge a fee for registration, it saw a dramatic drop-off. And the gaps in the data are also and the peaks in the China data are also a symptom of availability of data, because we used to have data for the second levels under .cn, but that is no longer shared at the moment.

But for the remainder of countries in the region, we see nothing dramatic. It's just not growing very much, and I think we can infer that difficulties with universal acceptance are really holding back growth in the region and also generally. Thank you.

However, that's the miserable news, but the good news is that when you look at the script of IDNs, you see that it's a very accurate and consistent

indicator of the language that you will find on web content. What we do is we strip away the parking pages and we go to rich content pages, and we see that Cyrillic script will really be associated with Russian language primarily, Hangul with Korean and so on. So IDNs really do have this potential to play an incredibly valuable role in enabling users to find content in languages that they can read and understand. Next slide, please.

We would like to talk to you about your IDN experiences. Members of my team here, Fiona and Georgia, would love to talk to people from the region who have experience in rolling out IDNs or other researchers or anything. Please get in touch. We always love to talk about IDNs with people. So, thank you very much for your attention today. I'm happy to take any questions either now or in the general flow. So I shall hand over to the chair now. Thank you very much.

SATISH BABU: Thank you very much, Emily, for that very interesting presentation. You brought out several very thought-provoking comments and very interesting for our region. We have some questions that have been asked in the chat, but we will take questions together after Sarmad's and Justine's presentation. So please bear with us.

So we now move on to Sarmad for the next presentation. Sarmad, over to you. You have 30 minutes.

SARMAD HUSSAIN: Thank you, Satish, and thank you, APRALO team, for inviting me to talk about IDNs. Next slide, please. So I will take this opportunity to give you a bit of background on the way they have evolved, and some of the challenges which came across and how we've addressed them and where we are as far as current policy development processes are concerned. And finally, I'll close with a bit of a discussion on universal acceptance. So let's get started. Next slide, please.

So basically, I guess early on in the history of Internet, mostly towards end of 1990s, the community had already realized that the domain name system which was at that time based on ASCII, as Emily was also sharing earlier, needed to be in local languages for Internet to properly proliferate across globally, I guess.

There was work already started in communities in late 1990s, early 2000s on looking at how to implement domain name system- in local languages. One of the early works was the work taken up by Internet Engineering Taskforce or IETF which meant we developed a standard for IDNs implementation and was published in 2003 as IDNA 2003 standard.

In the meantime, work also started at ICANN Org, and the community registries, both in the gTLD space and ccTLD space, actually came together and developed what is now referred to as IDN implementation guidelines. Version one of the guidelines was actually published in 2003 and these were guidelines which were focused on implementing IDNs at the second level, and they were endorsed by the ICANN Board and that's basically using those guidelines, the IDN implementation actually started under the gTLDs.

These guidelines have been part of the process as far as IDN implementation has been concerned for gTLDs. They're also recommended for ccTLDs who are implementing IDNs at the second level, and these guidelines, current, applicable version of these guidelines is 3.0 which was amended in 2011 and the latest one which was updated in 2018 is currently being considered by GNSO and will eventually, I guess, be taken up because the ICANN Board. Next slide, please.

So basically, beyond IDN implementations at the second level, there was also work which started to localize or internationalize the top-level domain so that the entire domain name could be in a local language, not just the second-level. One of the earlier works was done by the ccNSO within the ICANN community, and it resulted in what is now referred to as the IDN ccTLD fast track process which was approved by the ICANN Board in 2009, and that opened the way to allow us to have country codes in local languages and scripts.

This now has been in implementation since 2009. We saw the first IDN ccTLDs in 2020, and there have been some updates to this IDN ccTLD fast track process since 2009. I believe this version was published in 2019 with some revisions around string similarity review process. Next slide, please.

And this is a summary of where we are with IDN ccTLDs. We've successfully evaluated 62 applications from 43 countries and territories. Next slide, please.

Similarly, in the gTLD space, GNSO, as you all may know, worked on developing policy for new gTLDs, and the final report which was published in 2017 basically said that some of the gTLDs should actually be IDNs. That was basically the start of IDN support in gTLDs. And one of the things which was recommended in that policy as is listed as point 18 is that if an applicant offers IDN service at the second level, IDN guidelines should be followed. So IDN guidelines obviously continue to be part of the gTLD policy. And there are more details obviously available in the applicant guidebook which was eventually developed to implement this policy. Next slide, please.

To date, 93 IDN gTLDs have actually been delegated, and you can find the complete list online. And as you can see, it covers multiple languages and scripts. Next slide, please.

When the discussions around IDNs started, another thing which came up which is relevant to IDNs is that many communities, or at least a couple to start with, noted that some strings which cane be technically distinct were considered “same” by those script communities. So to them, even though technically they were distinct strings, they would want to consider them as the same domain label. Though it was identified that this was a requirement by the community, there was no clear understanding or definition of what this sameness actually means.

So these are normally referred to as variant labels, and they're relevant for two reasons. One of course is the security issue, because if they're considered the same by the community, then if these domain names are operated by two different registrants or TLD operators if they're at the top level, then you could imagine that end users would get very

confused because they may actually be going to the wrong websites in the best case. And in the worst case, somebody can use that to their advantage and actually get information from end users and steal their information. So there could be some phishing attempts.

And I've shared some examples to sort of explain why it's not that easy to explain what same means. On one hand, if you look at the Latin and Cyrillic example, [inaudible] in Latin, and similar, visually identical string in Cyrillic—and I've also put down the Unicode codepoint numbers. So technically, obviously, software is looking at Unicode and not really the string. It would find those two labels totally distinct from each other. But from a human perceptual point of view, those two strings, at least I can't tell the difference.

So that's one example. In the case of Chinese, I can actually as a nonreader of Chinese script, I can tell the difference, but the way it works in Chinese is there are two versions of the script, there's a traditional version and a simplified version and users, readers of the script would want to consider both as equivalent forms, so they would still want them to be the same. And the two scripts are used in different geographies. So there is a reason why both are needed at the same time and one should not be blocking the other. And at the bottom, you'll see example from Arabic script where there's some difference, but again, these variations come from geographically different locales and therefore both variants are needed for use and they may or may not look exactly the same. Next slide, please.

So basically, at that time, we're talking about 2000s. We understood that we as a community needed variants, not only second-level but also

gTLDs and ccTLDs. But it wasn't really clear what variants were, how would we define them, how would you use them. And it's very interesting that initial policies which came out actually knew that this is something that needs to be resolved and there were placeholders in those policies, but there were no answers, they were more questions. Next slide, please.

If you look at the IDN ccTLD fast track process, you will see that there's a placeholder there. For example, if you read the last bit, you would see that it says that the community is expected to continue working on more clear definitions on variant labels and solutions, or methods for delegation of a variant. So the fast track was not able to provide an answer on how to deal with variant TLDs. So that was an open question. Next slide, please.

Similarly, the new gTLD round from 2012-2013, also in the applicant guidebook noted the same thing that when a variant delegation process is established, applicants may be required to submit additional information such as implementation details for variant TLD management mechanisms and mainly to participate in a subsequent evaluation process. So they'd also in the gTLD [inaudible] put obviously the variant TLDs on a hold, because again, as I said, we didn't know enough at that time to solve that problem even though we knew that there was something which needed to be done. Next slide, please.

And again, knowing this issue and knowing the challenge but not the solution at that time, the ICANN Board in 2010 had resolved that no variant gTLDs will be delegated to the new gTLD program until

appropriate variant management mechanisms are developed. So we were looking for variant management solutions.

And following 2010, there was actually work started by ICANN community which looked at this in much more detail, but there were some variant issues reports which were done by different communities, and then from those multiple reports, there was an integrated issues report which was developed, and that issues report identified two problems which had to be solved. One noted that there's no consistent definition of what an IDN variant TLD is or IDN TLD is.

The second was that there's no IDN variant TLD management mechanism. So first, we needed to know what an IDN variant TLD is, and once we know, then we would need to know how to operate it or manage it. So those were two follow-up areas which were identified to address. Next slide, please.

So for the first problem, the community developed what is called the root zone label generation tools procedure. The procedure was developed in consultation with the community and the Board accepted the LGR procedure and asked ICANN and the community to go ahead and implement it so that variant TLDs actually can be defined or identified. Next slide, please.

So the way the procedure was defined, it was realized of course that ICANN doesn't have expertise to answer all these questions around scripts and how scripts can be used for top-level domains. So the expertise obviously resided in the communities which use that script. So a process was defined in which we were to support and develop with

the community panels for each script, and those community panels develop the rules and tell basically what would be a good, valid top-level domain and also what would be the variants of those top-level domains. And there was obviously an integration panel which would take an input from each of the generation panels for each of the scripts and integrate it into one set of rules which can then be applied to the root zone.

And that integrated set of rules which contained a list of permissible characters or codepoints, the variant labels and any other rules or constraints on labels, they could then be applied to any label in the root zone and we can identify the variant labels and the valid labels for TLDs. So this was a community-driven effort which started in 2013. Next slide, please.

So just to give you a current status of this, we identified in at least the current phase 28 scripts which are currently popularly used globally and used to write most of the languages around the world. And there was the LGR process identified some of the technical constraints which needed to be there. For example, adherence to IDNA 2008 standard and so on. And each community formulated their own panels, and you can see that to date, most of the panels have finished their work and the work has been integrated into the root zone LGR. We still have a few panels which are finalizing their work, and we have a couple of communities who still have to start this process. But the work on root zone LGR is largely now concluding, we should be done with at least the current phase within this calendar year, hopefully. Next slide, please.

Just to give you an example, here is the Unicode table for Arabic script, and there are some codepoints which are just not allowed for domain

names. Those are in white color. There are some others which are not allowed for top-level domains. Those are in pink color. For example, digits are not allowed at the top level.

And then the rest all were potentially gamed, and the Arabic script panel worked for multiple years, looked at what should be supported at the top level and how, and they have then shortlisted those which are listed in green color to say that these are the characters which should be allowed for top-level domains and the rest wouldn't be allowed. Obviously, if over time things change, Arabic generation panel can come together again and update its proposal. But that's what we have at this time. Next slide, please.

Similarly, the whole set of script communities have gone to the same cycle. So if we really look at it from a larger picture, this is I think from Unicode version 11 and covers 148 scripts and 137,000 total characters, of which IDNA 2008 supports 127 scripts and shortlists 97,000, almost 100,000 characters which are possible to use in domain name. Those left out include punctuation marks, symbols and things which are not allowed otherwise in domain names.

Current solution for example, the latest version of root zone LGR version four currently has a solution for 18 scripts which are currently used popularly to write most of the languages globally. And it shortlists from those 100,000 potential candidates 21,000 characters which can be used for top-level domains. So as you can see, the space for top-level domains much more constrained than the baseline which is offered by the IDNA 2008 standard. Next slide, please.

And then while these communities were working on shortlisting a solution for top-level domains, they also started looking at variants. And very interestingly, even though we thought it was a localized challenge, at least initially, there was just initial talk about variants by the Chinese community, by the Arabic script community, but when we started looking at it beyond these scripts, we really found that almost a majority of these scripts actually have variants.

Some are in-script variants, meaning there are two characters which are considered the same within the same script, but in some cases, there are cross-script variants where some characters—in for example Latin script are considered the same as some characters in [inaudible] script of Greek script. So there are both potentially cross-script or in-script variants. But there are variants in majority of the scripts. So that's at least what we have fond out based on the proposals we have received from the different communities. Next slide, please.

That sort of solves the first problem. The second problem was, how do we manage these variants which are generated? And for those, basically, what ICANN Org did was it actually undertook a series of studies and developed a set of variant TLD recommendations which focused more on the management of variant TLDs, application and management of variant TLDs.

ICANN Board approved these recommendations and requested both ccNSO and GNSO to take these recommendations into account when they started developing their policies to define and manage IDN variant TLDs. If you recall, as I shared, the current policies of Board, ccTLDs, for

ccTLDs, the fast track process and gTLDs, they have placeholders for variant TLDs but no solutions proposed.

So this provides a possible mechanism of both GNSO and ccNSO to consider. And another thing which the Board noted or requested that ccNSO and GNSO should really coordinate this policy development process, their own policy development processes across so that as far as variant TLDs are concerned, we can get a consistent solution across ccTLDs and gTLDs for a consistent, obviously, use by end users. Next slide, please.

So basically, as far as the recommendations were concerned, there were nine recommendations which were developed and published by ICANN Org. They are not technical recommendations because technical community has looked at IDN variant TLDs and they've really said that there is technically no clear solution. So I think the solution which has been presented is mostly around policy and administrative means on how to manage variant TLDs.

I'm not going to go through all these recommendations. You can go and look at the documents. But just as an example, one of the recommendations which is recommendation two basically says that if there are two different variant TLDs, they should be managed by the same entity. And that's obviously—the motivation behind that is since they're the "same" string as perceived by the community, it shouldn't be managed by two different people because that can potentially create problems for end users.

So there's these kinds of recommendations. And in addition to these recommendations, there are detailed analysis documents which have been presented on how these recommendations can impact the existing application and operational processes of gTLDs and ccTLDs. And so basically, the documents which were published by ICANN Org and requested by the Board for consideration by ccNSO and GNSO contain two pieces: the recommendation and analysis of these recommendations of the processes and procedures and application processes as well. Next slide, please.

In addition, since the root zone LGR becomes a very critical piece of this whole puzzle because it is essential to define, validate top-level domains as well as identifying, becoming the single source of identifying variant TLDs for any script, the Board also asks the technical community and all the SOs and ACs to look at root zone LGR and make recommendations on technical use of root zone LGR. So there was a Technical Study Group which was formed which also came up with recommendations on technical utilization on root zone LGR. These were also approved by the Board, and again forwarded to ccNSO and GNSO to take these recommendations into account as well through their policy development process. Next slide, please.

Basically, SubPro took these recommendations onboard, they considered these recommendations, the SubPro working group, and they actually have a whole section 25 on IDNs and a section on universal acceptance as well as some additional recommendations and implementation guidance which address many of the IDN variant TLD recommendations as well as were developed. Next slide, please.

In addition, as I shared earlier, the reports which were published by ICANN contained two pieces: the recommendations and the analysis of the recommendations as well. SubPro had looked at the recommendations and GNSO then formulated another scoping team to see how can some of the other detailed analyses that were presented in those reports can be taken onboard and reviewed and addressed as well.

So the scoping team basically suggested that a follow-up IDN policy development process should be started to address that, and they were also considering some of the new input coming in IDN guidelines, so the policy development process also wanted to look at how IDN guidelines should be updated from time to time and some of the additional items which are included in IDN guidelines for the second level. Next slide, please.

Basically, this summarizes the scope of what SubPro working group has done and you can see that they've looked at—so we're talking about existing gTLDs and future gTLD applications. SubPro is more focused on future gTLD applications and they've looked at from an IDN perspective, IDN implementation and IDN variant TLD recommendations.

In addition to that, as I shared, we have the analysis of these IDN variant TLD applications, recommendations from the Technical Study Group as well as some discussions around IDN guidelines, and that needs to be done obviously for both the existing gTLDs and future gTLD applications. And with IDN EPDP, expedited policy development process is being designed, the current charter is going to look across all these different aspects which look at IDNs for example in a much more holistic way

beyond what is already covered by SubPro working group. Next slide, please.

So again, as far as IDN EPDP is concerned, the charter for the EPDP has already been developed, and it is going to look at all gTLDs, current and future and the management of the variant labels under these gTLDs or of these gTLDs as well as the IDN implementation guidelines. And you can go to the link of the charter to seem ore details. Next slide, please.

And this goes into a little more detail on what it actually contains. I'm not going to go through all these details. These are different sections within the charter and all these sections have multiple questions around these different topics. So this just gives you a summary of what the charter is going to be looking at. Next slide, please.

The IDN EPDP charter has been adopted by the GNSO Council and the work is going to be starting. They've already called for volunteers to participate. I think that call is now closed, and the work is now starting imminently. Next slide, please.

So before we close, I also wanted to share a bit on universal acceptance, and Emily has already talked about some of the challenges. I just want to, I guess, categorize that a bit further. When we talk about universal acceptance, the aspiration is that all domain names and e-mail addresses are accepted by all the different applications. When we say all different domain names and e-mail addresses, these can be categorized by different types. So they can be those which are longer than the regular two, three or four characters, like .africa or .engineering or .technology. There can be new ones like .sky, there can be IDNs which

are totally in a local language, and then you can formulate e-mail addresses using these domain names of different kinds with either ASCII mailbox names or local language mailbox names.

And this creates many types of different addresses, domain names, email addresses which we normally test with different applications. And eventually, the idea is that the application should be able to accept, validate, process, store and display such domain names and e-mail addresses. Next slide, please.

I think we skipped a couple of slides. So just to give you a hint of where we are or how good/bad the situation is, what we did was we just went to top 1000 websites and in the Contact Us page, we just put in one of the different types of e-mail addresses which are listed on the left-hand side here to see whether those e-mail addresses are just accepted—not even checking processing of those e-mail addresses.

And we found out that for example for e-mail addresses which are purely Arabic or Chinese, of those top 1000 global websites, only 10-11% of those websites actually accept these valid e-mail addresses as valid. They think that these are invalid e-mail addresses. So 90% of these. So if you look at it, about 900 out of 1000 websites are rejecting these e-mail addresses as invalid, and that's obviously not a very useful situation. Next slide, please.

We also in another study looked at how many of the e-mail servers which are configured across the globe are accepting, if I send an e-mail to you from my Chinese or Arabic e-mail address and you're using a mail server, would it accept it or reject it as an invalid e-mail? And we found

out that only 9.7% of webservers globally in the sample which we took of mail servers deployed are configured to accept such e-mail addresses, and more than 90% would reject such email addresses. Next slide, please.

And you can actually check your own mail server. So if you go to UASG.tech/eai-check and put in your e-mail address, it will tell you whether your mail sever is configured to receive e-mails from, for example, a Chinese e-mail address or not. So you can actually test it right now. Next slide, please.

Obviously, this is a big problem. We want to address this problem. And to me, this has been one of the focus areas for not only ICANN but also obviously the community. Community has been leading this effort through Universal Acceptance Steering Group which is organizing the community to find out what the problems are and try to fix those problems. The UASG is developing multiple technical and other outreach documents to asses and address these problems and obviously, the community and ICANN Org are working together to do trainings across the globe to create more awareness and address this challenge. Next slide, please.

And that is all from me. If you have any questions, you can contact us at idnprogram@icann.org. Thank you. Back to you, Satish.

SATISH BABU: Thanks very much, Sarmad. Thanks for that very detailed presentation that brings up all the pending issues that we have to address through the EPDP. I see that Emily has answered her questions on chat, so thank

you very much, Emily. But please stay with us, there could be some more questions coming up. Any questions for Sarmad may be [inaudible] the chat box. And I note that we're running a little bit behind on time, so I'd like to straight away go to Justine for her presentation which is for about ten minutes. Over to you, Justine.

JUSTINE CHEW: Thank you, Satish. It's a bit hard to top what Emily and Sarmad have both covered. But thankfully, my role here today is limited to just subsequent procedures, and it's actually a short portion, really. Next slide, please.

Just by way of a little bit of background, what is subsequent procedures? Next slide, please. The way I like to describe them is think of it as the rules and procedures governing the applications for new gTLDs in the next round. They obviously relate to the new gTLD program, and because they only apply to new applications for new gTLDs, they don't apply to existing gTLDs. And this is borne out of the GNSO subsequent procedures PDP working group final report, the work that has been done by that particular PDP working group, and Sarmad has already alluded to that the final report is the last version of it anyways, that would be 2021.

It has basically what the GNSO Council have called SubPro outputs. SubPro is obviously the short form of subsequent procedures. And SubPro outputs cover policy affirmations, whether the affirmations are with or without modifications. They also cover recommendations and implementation guidance, and the vast volume of work that was done

by this PDP working group is just incredible because it covers 41 topics, and IDNs is only one of the 41 topics.

But essentially, these SubPro outputs are currently being considered by the ICANN Board, so it's gone through the GNSO Council. The GNSO Council has adopted it. And GNSO being the manager of the policy development process for new gTLDs. So GNSO Council has submitted the SubPro outputs to the Board for their consideration, and once the Board has considered it and in the event all the SubPro outputs are accepted by the Board, then we can expect them to be implemented by ICANN Org. Next slide, please.

Just touching on IDNs which is topic 25—and I've noted down the actual pages in the final SubPro policy development process working group [inaudible] if you're interested in looking through them. it's only five pages, really. While the SubPro PDP working group has affirmed IDNs as an integral part of the new gTLD program, in fact, IDNs are recognized as one of three string types, the other two being variant TLDs and geographical names. And that's pursuant to recommendation 4.1.

But the outputs per se don't go as far as to recommend priorities for IDNs, and that's through no consensus, really. But the outputs do recommend some prioritization in terms of processing applications for new IDN new gTLDs, and that's covered by recommendation 9.1 which is in a different topic, really. And one of the outputs also recommends to have the ability to accept both ASCII and non-ASCII characters in the applicant-facing system. So when the next round opens, presumably it's going to be a webform for people to apply through, and the idea is that you can then apply for IDN type TLDs by using known ASCII characters.

And in terms of the generation of TLDs and the variant labels, they're still subject to compliance and checking. That's recommendation 25.2. And some of the protocols and rules that you have to comply with are things like the root zone label generation rules. Those are used to determine which labels are to be blocked or which are allocatable.

They also need to be in compliance with the IDNA 2008 which is the protocols that I think Sarmad has mentioned. In fact, they are in reference to the IETF RFC 8590 to 8595. And in terms of usage of technologies, the outputs also suggest that where feasible, there should be a use of algorithmic checking for compliance.

And obviously, in the event root zone label generation rule has not been integrated yet—and we saw through Sarmad's presentation that there was about five that weren't still integrated—then the recommendation is to still allow for the application, have the application processed, but do not go forward with contracting or delegation and subject them to further evaluation. This is implementation guidance 25.3.

There was also a recommendation pertaining to single character in the sense that things like ideographs or ideograms can be allowed, and it would be allowed for limited script-language combinations. That's recommendation 25.4. But this is a conservative approach that we are taking in that any allowance of single-character gTLDs must not introduce risk of confusion from commonplace similarities. This approach is consistent with the advice of the SSAC as well as the joint ccNSOGNSO IDN working group. Next slide, please.

The rest of the recommendations, which is 25.5, 25.6, 25.7 and 25.8 deal with the practice of what I call keeping labels and variants together. I think there's been the term used to describe this, which is bundling. I'm not sure whether we're still using that term, but in any case. And this is important, as Sarmad has already suggested, to maintain this practice in order to ensure that the security and stability of DNS is maintained.

So, what do we mean by this? Well, if we look at the example that's up on the screen—I've used the example of Macau. It's not a good example because it is actually a ccTLD, country code TLD for Macau, but if I can just draw your attention to the character on the right of these two scripts, my Chinese is a bit rusty but I do believe that the character on the right is actually door, but the top one is in simplified script, the lower one is in the traditional script.

So if you imagine, it's actually the same word. And in accordance with the root zone label generation rules for Chinese language, these are the same word. These two characters represent the same word, except one is in the traditional form and the other one is in the simplified form.

But in terms of how the computer reads it, they have different labels. And you can consider them, in a simplified way, as variants of each other. So you can imagine that the chances of confusion could be very high because they're potentially the same word. And if they're delegated and used independently of each other, then we could run into all sorts of confusing situations, and it will also impact on the security and stability of the DNS.

So if you want to think of it that way, it's a bit like giving the dot door TLD to two different entities. If one party takes the simplified version of the script and another party takes a traditional version of the script, then it's akin to giving two different parties the same TLD, in fact. So that's certainly going to be cause for confusion. And it's inconceivable if we go back to the English version, the English example that I provided. We wouldn't give .door to t wo different parties, whether it's registries or what. And therefore, we need to maintain this practice of keeping the labels and the variants together.

So at the top level, IDN gTLDs which are identified as variant TLDs of existing gTLDs will only be allowed if the labels satisfy two conditions, which is that they're allocated to the same entity, the registry because we're talking about the top level, and if and when they're delegated, they need to have the same backend registry service provider. That's just to make sure that one single party has technical control over the label and the variants. And obviously, this arrangement has to be provided for in the relevant registry agreements.

Now, if you imagine the concept that I was trying to describe at the top-level and extend it downwards to the second level, then we need to have some kind of same arrangements for second-level domains which are variants of each other. And that is what recommendation 25.6, 25.7 actually attempts to do, with recommendation 25.8 which says that any second-level labels derived from the recommendations 25.6 or 25.7 are not required to act, behave or perceive as identical.

Now, this is based on SSAC's point of view, and because they've actually said that second-level domains—there is actually no way of ensuring

that second-level domains behave the same. It's not technically feasible to do that in the DNS anyway. So therefore, there's this requirement that they don't have to act, behave or perceive identically to be deemed as variants of each other. So we just need to make sure that any TLD that is applied for would have to comply with the root zone label generation rules.

So that's pretty much the recommendations out of the SubPro outputs pertaining to IDNs, really. I think a lot of the work that needs to happen still would be undertaken by the EPDP on IDN.

So I'm going to stop there and leave a little bit of time for questions.

SATISH BABU: Thanks, Justine, for the presentation. It raises the different issues that the EPDP will have to handle. Most of the questions raised have been answered in chat, but there's one question that's outstanding. I'm not able to understand. This is from Kossi Amessinou. The question reads like this: When we talk to universal acceptance for people speaking our language, we don't have online. What does that mean? This is for Emily. So I think this means a broader question of languages that are not available online. So what is the action plan for such languages? But Emily can try to answer this in a minute or two.

EMILY TAYLOR: Thank you very much for raising the questions. In fact, it's an area that UNESCO has been quite active about over the years, is about capturing non-written cultures, capturing cultures, but also indigenous languages.

They had a year of indigenous languages last year which we supported with a special study. Maybe that's available on our IDN world report website. Maybe have a look at that. But the experience of indigenous language speakers and the support for their languages within the domain name system is, let's say, quite poor. However, there are some glimmers of hope, particularly in social media environments. This is not something I find myself saying very much, but the support from multiple language speakers in social media is really good, and particularly harnessing members of those communities to help create menu systems and end-to-end environments within those closed communities. And that sees an uptick in the use of the language, which is really pleasant. And we feel like, well, if we could do the same in IDNs, then we could also enhance the experience of the web for speakers of indigenous languages too.

SATISH BABU: Thanks very much, Emily, for that. Before I hand over to Lianna for closing this session, there was a question that Dr. Gopal had asked in chat about the technical aspects of IDNs. And Sarmad has already answered it, but I'd like to add on a point that the UA tech working group is going to [bring out besides] this documentation also code fragments that demonstrate working code so that programmers find it easier to use.

So I'd like to stop here given the fact that our time's almost up, and I'd like to thank the speakers and hand over the floor to the vice chair of APRALO, Lianna. Over to you, Lianna.

LIANNA GALSTYAN: Thank you very much, Satish. I'd like to take this chance and opportunity to very quickly talk about the Armenian IDN ccTLD, [inaudible]. We're marking the fifth year of our IDN. So far, we have done a lot of work to advance that IDN and to develop the local content especially and therefore this year, we worked closely with the technical community, also advancing the universal acceptance.

And so we will have a lot of updates on this in due time. And it's [inaudible] we have this webinar for APAC region as we have lots of IDNs in this region. [inaudible] and we truly value the multilingual Internet and opportunity to use [inaudible] languages in our domain names.

And I'd like to thank all of our speakers [inaudible] in this topic, they presented a detailed explanation of IDN topic, the background of all the processes and what is coming for the next [inaudible] gTLDs and for the ccTLDs probably that will happen again.

Once again, thank you very much. We put these training materials and presentations, everything will be available on the dedicated page, so whatever you have, you want to refer to it, you can come back to it and check those materials.

With this, thanks again, and that is the end. Satish, over to you.

SATISH BABU: Yeah, it's back to Gisella for formally closing the session.

GISELLA GRUBER: Thank you, Satish. Thank you to everyone who has joined today's session. We have had great attendance. The recordings and the chat will be posted on the Wiki agenda page as well as the transcription if you'd just allow for a few days. So I'd like to thank everyone. Enjoy the rest of your morning, afternoon or evening, wherever you may be in the world. Looking forward to another session soon. Take care, keep well and safe. Bye.

[END OF TRANSCRIPT]