HYDERABAD – Proliferation of Indic Computing: Need, Challenges and Solutions
Monday, November 07, 2016 – 09:00 to 12:15 IST
ICANN57 | Hyderabad, India

[VIJAYANANDA PRABHU]: …from an [SD card] consumption method to an online streaming or consume from the Internet method. And, as we spoke about, the Play Store is in English, so unless somebody helps you with that, you can't necessarily go and download an app. So the method of pushing an app down somebody's throat even if you have content in a local language, it just doesn't work. So you have an issue with discoverability.

Then, you have the issue of ease of use. So we help a lot of customers go local with their apps and websites. And most of the method has been what the globe did, how did you localize something? You took something that was built and conceived in a certain language, you went and changed all the tags and all the user interfaces into a different language, and the users would input and read data in the same language.

Things work beautiful in Europe but [do] things work beautifully in China, Korea, and Japan? Because for them, English does not exist, they just have one language. So it's a unilingual method of accessing information and data.

We are probably – we, maybe Africa, East Asia, the colonized world if you may – are the ones for whom the entry has happened in English and the consumers are waiting to try and do this in multiple languages. Guess what? There are no methods available that let you transcend this even from a purely B2C perspective.

I'm going to talk more about how governments are, the enterprises are trying to tackle this because a lot of the data that they have is not content. It's actually data, which is residing deep inside their servers that Google doesn't get access to. So they can't do it on the fly.

So unless the enterprise or the government decides to make it available in multiple languages, it is virtually impossible for some people to get to it in multiple languages, which brings us to the point of how do they do this in 11, 12, 25 Indian languages. Because how are you going to maintain databases in X number of languages? You already have 30,000 webpages, how are you going to do your content management? Your content management team knows only English, how do you make sure that the content is right and so on and so forth?

Our company builds actually tools which allows you to do dynamic data across multiple languages on the fly, and so that's the plug for everybody. But coming to the fundamental issue

that we found in terms of why you can't get access to the content that already existed. There's some [inaudible] content out there, but it's very hard to get to.

First issue is with the encoding standards. Most of the encoding standards that started with ASCII started with the single character method of encoding. Indian languages don't work on single characters. How you get to a particular character towards the end determines how it gets encoded and represented. I just walked you through that.

And also the same thing applies when it comes to a display. And then, how do you help ensure that input methods are relevant because as Divyashree has mentioned, a lot of us, which are literate in multiple languages can only type in English. How do we ensure that we communicate to others who we understand that they do in other languages but we can't write to them?

This is for an example. The left side is Unicode encoding for all of Devanagari, which about three languages in India use. But what you see on the right is all Hindi needs. So there are about a half of characters out there, which aren't really legal. They just got formed because that particular character was borrowed from another language and imprinted into Devanagari for Unicode.

If you actually go towards a set of four characters, you see there are two out of those four are illegal. This applies to other

languages as well where things which are very close to each other. Because what happens is when somebody is actually trying to type, he picks the wrong because it looks very close by.

And tomorrow, when you're trying to search for it or access it, you can't find it because it has been returned as you so you'll know, and it has no phonetic equivalence, it's not necessarily legal. So there is no way you will ever get this content searched and found tomorrow.

The other interesting thing is that the same character that you see, which is [inaudible] can be constructed in two different methods. It depends on how it got constructed, it's going to show up in completely two different search results.

The character at the bottom is again illegal. It can be constructed this way because the [matter] on top is what you will put over [inaudible], but that's not what that it's really supposed to be.

If you look on Google, we have searched for what appears to be the exact same search string, and it shows up two very different sets of results because how the input in the search was put in and you see that. What it basically means is that half of the content sitting out there is not really discoverable in the intent that it was meant for.

When it comes to the input from a native perspective, our keyboards typically cannot be a single layout keyboard. It has to be intuitive in a method that the first character. And then, once you do that, what you're allowed to do with that character comes up and you can actually allow only things which are legal and then it can [become]. So this is an example of something which was not done very correctly.

The other thing that happens is that since a lot of people are new to this, they type something while they actually mean something completely else. And this we see it manifesting even they don't have Indic keyboard and they still type in English. Understanding that the method what people are using to search and what they are typing are two completely different things, very different from how it's been happening in the rest.

We've been working on a lot of this, which tries to understand how the Indic user interacts with the Internet, how are they looking for things and so on and so forth. A lot of work has to go on because there is not enough [corpus] out there, which allows you to look at and form these things. It will happen over a period of time. But what we've seen working is that when people are actually doing this with voice, a lot of these problems kind of go away because they're actually saying what they're looking for.

This is what we found somebody looking for a mobile cover for one of our customers. This is what they actually typed. In the Eastern part of India, that's a very valid method of how they would pronounce mobile cover. That's how they speak ,so they just type phonetically and you will not find mobile cover if you do that.

The next thing that happens is Indians are inherently multilingual, 80 plus percent of Indians speak two languages at least, more than 40% speak three. So they interact continuously in multiple languages. This is the search that we have seen: half of it is in English, half of it is in Hindi. So these are things that need to be taken care of also for [discovery] to happen.

So that's primarily what I had from a presentation perspective.

UNIDENTIFIED MALE: Great. Thank you. I think on linguistic space, English and Hindi on Google works very well. I have tried several times, by mistake it happened, and it gives me the results.

We have [reached] the Q&A. Are we waiting for [inaudible]?

UNIDENTIFIED FEMALE: [Inaudible].

UNIDENTIFIED MALE: So let us have the Q&A. If there are any questions from the floor? Yes, please.

UNIDENTIFIED FEMALE: A couple of questions, a quick one, as an Indian, when would I be able to type the domain name in my language?

Two, when we come to questions where the language is an issue and interpretation in English is an issue, not everyone is well versed in English, would a search, a voice search help perhaps, which can actually – the system can understand? Google does have a voice search I understand, but can it interpret the way I pronounce or even a Hindi word or a local word?

And the third one is, when you talk about different keypads for typing in for local languages and even in multiple local languages in English in my mobile or in my laptop looking at the cost-conscious population you're dealing with, could you think of a single unified cost-effective platform where they could use it in a similar manner?

UNIDENTIFIED MALE: [Inaudible] would like to take it?

UNIDENTIFIED FEMALE: Yes. I'll take the second one. The first one, domain names I think Dr. [Data] just mentioned before.

UNIDENTIFIED FEMALE: Yeah.

UNIDENTIFIED MALE: You could answer all of them. But you [inaudible].

UNIDENTIFIED FEMALE: Then, please, I can talk about the voice search because voice search is designed to understand different accents. And if you [just start] the Hindi word search, you will see that it is designed to understand different accents. Basically how you even look at English, that's why we build the entire speech-to-text platform wherein we do test it out with different accents, different voices, different pronunciations.

Especially in India where we see that in our language like Hindi spoken in eight different states, and then in each state a lot of words become different, a lot of pronunciations become different, so it is definitely designed. We do take a lot of voice samples and build it on that. And you will see that in English, Hindi, or any of the voices that we support. Any of the voice

searches that we have in our system, it will definitely work with that. It's designed to [inaudible].

[AJAY DATA]: So actually, on Hindi, your voice and my voice is perfectly [understandable] by Google Voice today. I type my all Hindi [mails] by voice, and pretty much almost accurate and sometimes [small letters] because of the speed at which we speak or something like that, but it can be corrected with time, and this works very well.

Typing on IDN domain names, so if you are pure Hindi typing user, you can change the operating system language itself and add a keyboard of Hindi. Or if you are an English Hindi, then you can add a Google input add-on into your Chrome browser, so it gives you a [translation]. Chrome has a Google input add-on and you can have a [translation] on your address bar. And actually, that enables to type anywhere in the browser then. So even [one subject] will be used for all the browser bar, it will be useful.

Yeah?

UNIDENTIFIED FEMALE: My question was slightly elongated. When I took Hindi as an example there, officially, we have 22, we are working on 15, so I'm looking at the other languages too though Hindi is more

developed. But when we are talking about making India [reaching the masses] digital literacy and all that, how it [weaves] and what the government is also doing and what do you expect the government to do. That's possible.

There are a lot of things being done but there are certain gaps, too. Else, we would not be sitting in the room.

UNIDENTIFIED MALE: Government, we have ICANN to [commit] much. Operating system supports all the languages, and so as Google index supports I think 11 languages.

UNIDENTIFIED FEMALE: The keyboard supports 11 languages.

UNIDENTIFIED MALE: Eleven languages, Input Tool, which covers the 90% of the population with the masses. Obviously, we need all 22, so that probably Google must be working on that. And eight is good enough right now because there are only eight domain names in Indic languages [as of now], so you don't need a domain name to type in that, then otherwise you will require in the search bar. Domain names, you do not have an Indic more than eight. So eight is good enough right now.

Probably, yes, you can add, yes, [Reshmi] can add more.

[RESHMI HARIHARAN]: Actually, we have 15 domain names in India operational since August 2014. In Devanagari script, we have seven, then other languages. Totally, it comprises of seven scripts covering 15 languages. And we have recently applied to IDN ccTLD Fast Track Process of ICANN for eight more languages to cover the 23. So in another six to seven months, all the 22 domain names we would be having.

UNIDENTIFIED MALE: [Inaudible] where we have [inaudible].

[AJAY DATA]: You will have – yeah, so another six, seven months.

UNIDENTIFIED MALE: It had to begin with Hindi because that is where most of the traffic currently is sent from.

UNIDENTIFIED MALE: Traffic is sent.

UNIDENTIFIED MALE: If you look at the figures, which Facebook for example shares, about 70 million are Hindi. Then you look at Kannada, Marathi. Bengali seems very high, but I think a lot of Bengali is skewed because of the Bangladesh numbers which get added to this. Otherwise, if you look at Hindi, Marathi, and Tamil, as of now perhaps, the largest population on the Net right now is taken care of.

UNIDENTIFIED MALE: Yeah.

UNIDENTIFIED FEMALE: I had another question. Since you come from the content space, how are you looking at the usage, the local language usage in the Tier 1 and even from [CSCs] or the lower level cities? Are you seeing some traction happening there?

[RAJESH KALRA]: In fact, the growth is happening there. Initially, maximum growth was seen in [inaudible] times. But now, we see it in all other languages. In fact, last year, we started in three places, in three languages where [at the Times] it doesn't have a print presence. That is the Telugu, Tamil and Malayalam.

And this year, we're going to add two more to that. And Gujarati is already there which was a print presence but now it's – next three months [over the] transition, it will be only digital. So all these areas we are going in only because we see a lot of potential there and the traffic is coming from very, very small towns.

UNIDENTIFIED MALE: Yes?

KELVIN WONG: Hi, from the ICANN APAC Hub, not a question but just a comment. There's an initiative from the APAC Hub and now it's with the Language Services Team in ICANN on the Language Localization Toolkit where we have a deck of slides that explains about ICANN Universal Acceptance and IDNs, and it's free for the community to translate into different languages.

And we have actually that translated into Hindi and I got reminded of that because we have somebody in the room who actually did a great part in that. If you like to just raise your hand, [inaudible]. Thank you so much.

Yeah, so this material is available in Hindi and it's really from the community. So it has been translated in the various languages as well. I just wanted to point that out.

UNIDENTIFIED MALE: Very good. So we can have the entire ICANN site in Hindi?

KELVIN WONG: This is really a deck of slides that really explains ICANN in very simple terms. It's really for outreach purposes that you can bring to the wider community but it's not the entire ICANN site.

UNIDENTIFIED MALE: So this is actually – I wanted to put on record that when we are promoting this IDNs and everything, so ICANN should be the first one to convert this entire language into other languages and make them available on the IDNs. So that we become the – [inaudible] measure for that. Because otherwise we are only preaching, we are not practicing.

DON HOLLANDER: Hello, from New Zealand – a question about the Times of India and your websites. And you'll have some place where people can register with their e-mail address to receive information or to be able to access information. Where people enter their e-mail address, can they put in a [inaudible] e-mail address?

UNIDENTIFIED MALE: Sorry?

DON HOLLANDER: Can they put a local language e-mail address in the Times of India application?

[RAJESH KALRA]: Local language address, not as yet.

DON HOLLANDER: But you'll fix that shortly?

[RAJESH KALRA]: We will fix everything that is required to get more users.

UNIDENTIFIED FEMALE: Don, if I may point out, the local language [inaudible] does work with the Times group mail client.

DON HOLLANDER: It does?

UNIDENTIFIED FEMALE: We tried yesterday.

[RAJESH KALRA]: She corrects me. That's why she's my colleague.

DON HOLLANDER: So that was the e-mail system?

UNIDENTIFIED FEMALE: Yes.

DON HOLLANDER: But it's your subscription system.

UNIDENTIFIED FEMALE: And this is the Internet [inaudible] wouldn't work with that?

[RAJESH KALRA]: No, it doesn't.

UNIDENTIFIED FEMALE: Okay.

[RAJESH KALRA]: No, as I said, if it is required, we'll do it.

UNIDENTIFIED FEMALE: Okay, there you go.

UNIDENTIFIED MALE: Yeah? Yes?

UNIDENTIFIED MALE: [inaudible], my question is about the times that since we are going into the more of the discoveries and those [can speak] kind of thing, what about the e-commerce site? E-commerce site where the most of the business is lying in the [Type 2, Type 3] cities like Flipkart, Snapdeal, and travel portals? What about the response of these portals, the service, the content development in the local languages?

UNIDENTIFIED MALE: See, this is an e-commerce [field]. I'm more from the content side, but I can tell you that the behavior all over the world is almost the same. India is slightly more complex because, as he was mentioning, the accents are different.

In fact, if you go to Punjab, you go to different parts of Punjab, people speak Punjabi in different ways. I am a Punjabi, but I don't understand 80% of the things which some other parts of the state, the way they talk.

It's a challenge and they're all trying to see what can be done to make it easier. [But as I] said if there is a customer there who wants to place an order in a particular way, even if it is 1 out of 200, he will make sure that at least that is made available here

because everything which is possible technologically, it can be built into the system and made available to the person.

The only issue then is, as we realize with our own apps and everything else, every time you add a feature, it makes your app heavier.

UNIDENTIFIED MALE: Can we have a last question? Yeah, [you want to ask?]. Yeah.

NEHA ALAWADHI: I'm from the Economic Times. I want to check with Divya, you've mentioned something about a handwriting tool.

DIVYASHREE BHAT: Yes.

NEHA ALAWADHI: The first question would be, does it support Indic languages? Does it provide that kind of support? And this is [geared] more from server request to all review panelists. If you could just narrow down or at least put out, because there's a bunch of challenges that you've mentioned and all of them are valid. Are there focused areas from this session if I could take back and say that these are the focus areas that the industry is working on? Could you maybe put that out when you conclude the session?

DIVYASHREE BHAT: To begin with, yeah, the handwriting input app, it supports Indian languages. In fact, post this session, I can run a quick demo for you. I have the app and I use it all the time to actually communicate in Kannada, and it works very well. As a user, I can tell you that it works really well.

In terms of the challenges in what we're working on, it comes back pretty much to content and awareness, which literally is for us to be from what we're solving for it, Google. It comes down to how do we really look at growing in the language content, and then how do we enable users to communicate a lot more in Indian languages.

So you see that we've started putting in a lot of effort on building awareness around Hindi search, improving Hindi search experience for the users. So that would be our big focus right now. At least, you start getting Hindi right, and then we can expand it to other languages far more easily.

RESHMI HARIHARAN: I just have a question to the audience. Since different companies are working on developing different keyboards and different organizations are developing operating systems in multilingual languages, there is this issue of transliteration where the

standard terms might be represented differently in different operating systems.

From the government, we had been receiving proposals from different companies like this is our proposal to standardize the terms. I would like to know whether the industry is collaborating to get these standard terms out, or should the government take the initiative and handhold all the industries to get it in the first place, so that when keyboards and operating systems get translated into different languages, there is some uniformity? Comments are welcome.

UNIDENTIFIED MALE: Yep?

ANIVAR ARAVIND: I'm from Indic Project for the Malayalam Computing. Yeah, [inaudible] I actually need to asking three questions. One is on the side note circulated by [inaudible], it looks like the whole Indian language infrastructure problem is listed as a part of [content] because these are two separate problems. Infrastructure need to be treated separately.

And also, as a part of that, the second question coming in, which is last 15 years of state investment in the language infrastructure development, which is not still available to public or not

available as a basic infrastructure for [startup] for community to [work on].

When we talk about the discoverability problem, I think we are easily going to the solution of maybe appropriate research or property [inaudible] solutions proposed by Google at the same time. I think why we are unable to develop maybe a cross-language search or a [inaudible] search – a cross-language search or approximate search, which will be usable for in the languages.

And it also comes to the question of why, that all the state investment came into this domain, it's not released there until – under [inaudible] [institutions]. So without the public access into this under a free and open [inaudible], I think the way forward is a bit difficult.

And the third question is, even in this problem because infrastructure problem, I think that there are many areas just like standards and [coding] even [inaudible], stemming. And on all this domain and including [sorting] orders collection and everything.

So I think in these areas, as of now only open source communities are working in India. At the same time, this technology domain there is no support available. And as of now in India, the technology development or [Indian] languages is

just government domain or government-funded [inaudible] domain [inaudible].

I think the way forward is more like a multi-stakeholder approach in this domain, including the open source communities, the civil society as well as [inaudible] even. [inaudible] if you look at the Indian language technology, I think there are a lot of open source communities active in India that need to be accommodated because Indic language industry is not much active in the technology infrastructure field. Because then you can avoid those issues like separating content and infrastructure.

UNIDENTIFIED MALE: Good additions. I think—may I should note that and we should move to the next session.

ANIVAR ARAVIND: Just one more point because apart from Google, there is one keyboard called Indic Keyboard. Indic Keyboard supports 23 languages and 54 layouts in that, which is much bigger than a Google [inaudible]. It's totally open source and also compatible to the Android open source.

UNIDENTIFIED MALE: Great. Perfect. We are done with the questions. I have been told to wrap it up.

UNIDENTIFIED MALE: I do have a question [inaudible].

UNIDENTIFIED MALE: Okay.

UNIDENTIFIED MALE: Yeah. [Inaudible].

UNIDENTIFIED MALE: [Inaudible], yeah.

AKSHAT JOSHI: This is Akshat Joshi from C-DAC. I just wanted to respond to the question posed by Reshmi regarding having standard terms for those terms used in computer systems or digital systems at large.

We, C-DAC, along with [inaudible], we have already started a project a few years back that is called the [fuel] project that is frequently used entry list. It is specifically focused towards standardizing by way of involving community into getting the standard terms for all the digital terms that are used in

softwares like file, open, save, and there are many other things. If you go to the website, you can see that.

I would like to use this forum also to request everybody to go through that and please join if you can.

UNIDENTIFIED MALE: Thank you very much. Thank you for the participation.

End of this first session and Don, you might like to take over for the next session.

DON HOLLANDER: Thanks very much. So if I could get [inaudible] and [inaudible] to come closer to a microphone, please.

UNIDENTIFIED MALE: Don, may I ask one question while we're changing the room? To the gentleman who just spoke, I'm wondering is your project tracked anywhere in something like an RFC process? I think it's good to memorialize these projects, basically give them a number and put them into the process. Everyone can create an RFC, and you've done a lot of work. This would bring it into the mainstream I think and allow you to get more comprehensive feedback perhaps.

UNIDENTIFIED MALE: Yeah. I don't know exactly where it is being tracked, but I can assure you that since [inaudible] and [inaudible] are involved, there must be some kind of writing, so we will get it offline. I'll give you the coordinates.

UNIDENTIFIED MALE: Thank you. Thanks.

RAM MOHAN: Yeah. Just a quick follow-up from what Mark was saying. There might be an opportunity to take the work that we're doing here in India and put that into an IETF RFC type of a process so that these standards that are being developed here become adopted and become kind of open standards available for the whole world.

Because if you're not from the India and tracking what was going inside India, it is hard to find out precisely the technical details of the work that you're doing.

UNIDENTIFIED MALE: Thanks for that solution. I'll take that to the field project. Thanks.

DON HOLLANDER: I'm the Secretary General of the Universal Acceptance Steering Group. In this next session, we're going to talk about Universal Acceptance: what it is and why it's important and how easy it is to implement into an application.

Universal Acceptance is the idea that all computer applications will accept all domain names and all e-mail addresses equally. When I asked question about the Times of India and the ability to log into the account and the ability to use an Indic language e-mail address, and the answer was, "Oh, probably not, But if it will generate more users, that's something that we'll do," that's a generally correct answer that right now, most applications will not work with the IDNs, the Indic language domain names and e-mail addresses, Arabic names.

So what has happened? Just two sentences. What has happened is about 2010, the top-level domains, the domain name space started changing quite rapidly with the introduction of IDN domain names. So IDNs are internationalized domain names. Though, they're not internationalized if you're in India, are they?

These are non-English or non-ASCII domain names. There are about 30 to 40 ccTLD names. And this is because the communities outside of North America and Europe wanted to have the Internet in their language. It was a very big issue for people of the Asia Pacific region, and they created a program

called the IDN ccTLD Fast Track, which we heard referenced to earlier.

So now, you can have domain names, the complete domain name in your native script. And then about three years ago, 2013, there was a flood of new top-level domains coming in, over 1,000 to 1,200 new domain names, some in non-ASCII characters, non-English characters, and some in English characters but new ones.

So .plumber for example or .guru or .technology or .photography, those are all longer than your traditional two- or three- or four-character top-level domains. And some of those also don't work when you try to log into your local tax system or your local voting system or your local banking system. So it's not just people with IDNs that are currently constrained but these new domain names that are available.

The Universal Accepting Steering Group was created about 15 to 18 months ago to raise awareness of this to the CIOs, the computer programmers, the systems architects and get them to realize what's involved to making them happen.

I've asked a number of people today to come and talk to us about their experiences in making things happen. These are real-world practitioners who get their hands or get their staff's hands dirty in actually doing the coding.

We have Ram Mohan to my left, and Ram is the Chief Technology Officer for Afilias, one of the world's largest registries. They also offer the backend registry service for .in and the associated ones there.

Dr. Ajay Data is the CEO of a number of IT companies in India. XgenPlus recently, about two or three weeks ago, launched an e-mail application both browser-based and app-based that allows unicode@unicode.unicode addresses. So totally non-English, non-ASCII addresses. They're also my provider for my arabic@arabic.arabic or actually it's arabic.arabic@arabic e-mail address. I'm very excited that I got that working just the other day, and I'm sharing it with people here and they're all jealous as can be.

Mark Svancarek works for Microsoft. He's driving their Universal Acceptance Program. He's also driving their IPv6 program and some other programs there. For someone so young, Mark has a vast experience in innovative technology. Mark is also the Co-Chair of the UASG's EAI efforts. I'm saying that and I'm saying there's an awful lot of acronyms, and I'm sorry.

Ashwin is the CIO of ICANN. ICANN has a vested interest in making sure that Universal Acceptance happens. It's not a commercial imperative, but it's a community imperative that

people can use whatever identity they want, whatever e-mail address they want to engage with ICANN.

And [Sigmund] is over there and he's driving the program for ICANN. So he's actually in the nitty-gritty of what's happened.

This is going to be a panel discussion, no slides. I know all of these people, and I think this is my last opportunity to actually say anything.

So I've got just a couple of questions for each and then if there's time left over, we'll have some discussion. So Ram first, why is UA important, particularly why is it important to Afilias, as well as generally?

RAM MOHAN: Thank you, Don. Let me tell you a story. A short while ago, about seven or eight months ago, I registered a domain name or I had use of a domain name in Tamil. Those of you who were at the public forum will know Tamil is my mother tongue. So I registered a domain name in Tamil, and I wanted it to be ram@mohan.india, which is a natural thing. I'm an Indian and Mohan is available as a domain name in .india. And I put my first name in front of it. That worked, or so I thought.

The entire process was quite effective, and one of the basic litmus tests I do these days is after I do such things, I come to the

ICANN site and I see can I register to the ICANN Conference using that e-mail address?

But I thought now ICANN is a global organization and they have lots of competing priorities and lots of things that they have to do. So I thought, let me do something far more direct. What's more likely? If I was a Tamilian in Tamil Nadu or whatever, I register a domain name, what are the things I'm likely to do as a Tamilian? I'm likely to go online and I'm likely to look at Tamil content.

I mean, that's a natural thing. So I can go to samachar.com or I can go to Vikatan and Vikatan online is one of the largest well known publishers. They have a strong presence in the Tamil content area.

Let me tell you, when I typed in my e-mail address in there, it didn't work. It is ironic because here are all the things that we talk about that are important for proliferation of Indic computing. We have Indic content. We have Indic e-mail, Indic top-level domain name, Indic second-level domain name. So from a checkbox point of view, every single box is checked. Yet, it doesn't work. And that in a nutshell is really why Universal Acceptance is the highest priority for us to focus on.

In many ways, the fact that you can get a domain name, you can get an e-mail in your local language, while it is important and

while it is part of the foundation, at the end of the day, if you really want to proliferate Indic computing and if you really want to have digital natives in local languages come forward, you have to solve and we have to solve the Universal Acceptance issue.

It's basically quite a strange phenomenon to spend money and time and effort to represent my own identity in my language, in my country and yet have my people, my programmers from my country not actually taking the steps to recognize these domain names, these e-mail addresses and take the right programming steps that are necessary to actually accept these names.

I find that quite a difficult thing to stomach but it really motivates me, and I hope it motivates you to get efforts going towards getting to Universal Acceptance. Because from my point of view, if we get operating systems that are in prevalent use here in India, if we can get the browsers that are normally used here, mobile systems that are in use here in India, if we can get Web validation forms.

Now, I went and I looked at the source code behind what was going on here. And to the best of my understanding, the coding was simply leveraging a library that somebody had downloaded from GitHub and had reused. So they had used the library and the library basically said, "What is a valid domain name or what

is a valid e-mail address?" And that library ended up not accepting my Tamil e-mail address.

After having looked at that, I thought, "Okay, maybe this is only a problem about IDNs." So one of the Indian companies that has a top-level domain name working is Dabur. All of you know Dabur. They're a huge company. They are very popular, etc. They have .dabur.

.dabur not in Hindi, not in any other – it's in ASCII, it's in English, Dabur. And I tried to register with an e-mail address in .dabur. Failed. And the failure was because the libraries or the programming – whoever had programmed had used some system that had a default that said the only valid top-level domains must be two or three characters in length.

I mean, this is like the equivalent of 19th century thinking, but that's where we are in the current contemporary situation here in India. And that is really why Universal Acceptance is so critical. The fact is that to actually get websites to accept domain names and e-mails that are legal, that are in local languages the technical work required is not very high.

But if I'm a programmer, and I had been a programmer before, I will say this. If somebody came to me, if I came and listen to me talking here and saying this, I would say, "Why would I bother?" I already have preexisting libraries that I'm using. I'm pulling this

down. Why would I want to go change it, modify it? I have deadlines to meet. And honestly, there is not much complaints coming through. Nobody is saying this is real problem.

So I think the big challenge is to go and ensure that the actual code libraries get more refined. The challenge and the job in front of us is to make sure that all of our businesses, all of the e-commerce sites that are here in India, content sites that are here in India, that they initiate a fundamental review of what they are doing to become UA-ready. Because becoming UA-ready is not a very difficult thing.

What we have found in our engagement around the world, and I think it should be easier here in India because they're such a technologically advanced country, the primary issue is not the technical challenge. The primary issue is to get a decision maker to recognize that you have to tell your programming staff, devote, it probably will take half a day. It probably will take in actual programming time, in total elapsed time, it might take a week or so. But that's about it in terms of what it takes to get some basic things UA-ready.

Once you get to some basic things like form validation, e-mail address acceptance, you get those two things ready on the UA side. Now, you have enabled people who are using Indic

language domain names to actually start to see some value for their domain names.

Once they start to see value, it is likely that I can write from ram@mohan.india to rajiv@bansal.bharat and have it work between each other's systems. Because technology is getting in the way and Universal Acceptance is really the project to help technology get out of the way for people to connect together.

DON HOLLANDER: Thank you very much. Ram talks about the various steps involved in getting UA-ready. We have a pile of these brochures, so we have printed brochures available in English. And on the website, we have them in available in about 11 different languages including Hindi. And if there's a desire for us to have them in other Indic languages, just let us know and we can do that, too. That's not a worry.

Mark has been working on this at Microsoft for a little while. Mark, how big a deal is this for Microsoft and what kind of efforts and how much effort has been deployed?

MARK SVANCAREK: You may be aware that Microsoft recently, a few years ago, got a new CEO. It was a pretty big deal, and a lot of cultural changes happened. When I joined the company, our mission statement

was put a PC on every desk and in every home. And about 15 years ago, we had accomplished that and we searched for a new vision and sometimes we had trouble with that.

Our current mission statement is enable every person on the planet to achieve more. So immediately you think using your own language in your Internet experience would allow people to achieve more. It's a large hurdle.

At a very high level, this resonates within our current culture. Now, we have a large portfolio of products and services, and large code bases, some of which contain a code that may be 20 years old. There are parts of Exchange that are very old. There's parts of the Office suite that are very old. As we need, we rewrite portions of them but if it works, you don't break it and open it up again. And so the experience of EAI and IDN was very variable.

Because we're a large company that is run under different executives, there was one portion of the company, the Server Group, who made it a requirement to support IPv6, to support fully qualified domain names, restful APIs, EAI, IDN. There's about 25 things.

Active Directory has always supported IDN and EAI, SQL server, Windows server, etc. But on other parts of the company, it was very hit or miss. The Office Team is always very forward-looking, so they had been implementing portions of this

opportunistically, but they had never committed to delivering end-to-end solution for EAI. IDN, most experiences already were enabled but even there, not all experiences.

So the first challenge was to go and convince the other parts of the company that they should submit to the rules that applied somewhere else. This is not the way Microsoft previously was run. Our new leadership is spending a lot of time making sure that we can work together and share code. Historically, we all used different libraries or different versions of libraries, different bug databases, etc. No one wanted to take a dependency on someone else who maybe would cut a feature or slip a schedule and stuff. And so that culture has had to change.

But it was still an interesting task convincing people that this was both the proper thing to do and an appropriate thing to do for the business. We really train our program managers and our engineers to be very skeptical about passionate claims by people who are enthusiastic about new and shiny things.

And so they would ask, “Well, what about biometrics? What about using mobile numbers as identity? Won’t those make e-mail addresses go away?” And so we would have to have these discussions.

We live within a bubble. The engineers in Redmond have all gone to good schools. They were raised in households that had

exposure to English and the Latin character set. Their parents may be conversant in English as well and so they think, "Well, this doesn't seem like it's a problem."

One of the resources that I had to use was a public service announcement video. You can see it on YouTube created by THNIC in Thailand called KaiKaiKaiKai. Mr. Kai wants to sell chicken eggs. And in the Thai language, which is tonal, it's not actually kai-kai-kai-kai. There are variations in those, but when you transliterate them to ASCII, that looks the same, kai-kai-kai-kai.

And unfortunately, that has about eight other meanings depending on how you – and so they show a little video how he gets his domain and nobody can figure out what it is he's doing and it's tragic. And when you show this to the engineers at Microsoft, they say, "Ha, now, I get it. Now, I see." This really is a big gap for us.

Once we had some buy-in both at a management level and at the grassroots level, because we do work top-down and bottom-up in a lot of ways, then we were able to start attacking the problems at their sources.

The first thing we went after was the Outlook desktop client for Windows. That has been EAI-enabled for quite some time now.

Then, we went after the Exchange Server; the Office 365

online servers, which is built on the Exchange Server; Office for Mac; Office for Android; Office for iOS. Those are all moving ahead nicely, and someday soon I will be able to announce that that we support this across this large part of our product and services portfolio.

We are still not able to offer the provisioning of EAI identities yet. In that, we are like Google. Google has not been able to offer them as well. So it's very exciting when you see GENPLUS and they offer creation of identities in – I think it was seven or eight – eight Indian languages, seven which are Indic, one is English and three other languages.

So now, more people can actually use this in a reasonable way and also now, we can do some real-world testing. I'm happy to say that my Office 365 tenant is working great with all of these, and we haven't actually announced it yet but it's out there right now. If you have Office 365, your mailbox probably already works.

Within the company itself, we had to think about what does it mean to be UA-ready. The line of business applications that we write internally, some of which again are pretty old, and also our relationships with our vendors, SAP and the like.

And so we are now just starting the process of scanning through our agreements and our software, and the like to find out exactly

how close we are to being UA-ready, similar to what ICANN has done. I anticipate it will be a similar effort to what ICANN has done, although there are more tools right now available to us. They were the pioneers, and we can stand on their shoulders as it were and save it ourselves some time.

However, since most of our internal applications are built on .net framework or the Universal Windows – this is sort of a pitch for our developer [story] – of the Universal Windows Applications platform, those are fundamentally EAI and IDN UA-ready. Anything that has been written using .net for .x, you just get that for free. But still, there is an auditing process that it goes in.

So in summary, we take this seriously from an ethical perspective as well as from a business opportunity perspective. Our engineers are now on board and we're chugging away, and we have great progress to report every quarter. And internally, we are now beginning our audit process to find out how UA-ready we may or may not be so that we can apply some pressure on our vendors as ICANN is doing to move the rest of industry ahead.

DON HOLLANDER: Thanks very much, Mark. There's a lot of bits in there, so thank you.

Ajay, your company gets recently decided or you certainly recently launched your EAI product, EAI solution. When did you make the decision to pursue it, why, and how did you approach it and what bumps have you found in the road?

AJAY DATA: Great. Thank you very much Mark for acknowledging [xGen]. We heard the government of India come out with .bharat in around 2014. We got an e-mail because we have a trademark named Data, which is our brand name. Because of some [grace period], we can register our own .bharat domain name, and that is how we came to know and we registered our .bharat by paying, I think, 4,000-odd rupees at that time. Welcome, [inaudible].

And at that time we paid – and we thought that who is going to use it, how are we going to use it, all kind of challenges. We have a domain name. We have paid money. But fortunately, we were an e-mail company, so we were providing around 11 million e-mail accounts in English for enterprises already.

We are an e-mail company and we have enterprise solutions, so probably we have the largest deployment in Asia around hosting one company with six million e-mail accounts on our platform itself.

So e-mail was a background last 15 to 17 years we have been working on, and we started looking at what will entail to build this capability of linguistic space into this e-mail. And it's a pretty tough task I must say because it is not just e-mail.

As Mark has pointed out a few points, you need AD, Active Directory. In our case, it was [database]. You need anti-spam. You need anti-virus. You need logging engine. You need SMTP. You need to also do on the fly that the recipients receive an e-mail or not, SMTP UTF-8 as we all know that.

So there are a lot of pieces at around there, then backup and restore, then preserving those files in that particular format. All those pieces have to talk to each other together. And I think this is probably one of the reasons a lot of companies are not able to do it faster because the components are too many. There are about 14 components.

We all know that on enterprises space, anybody who's offering Microsoft Exchange for example from Microsoft, do not offer anti-spam. So you need to have a third party anti-spam vendor to support that capability or anti-virus, same as for Google, same as for all enterprise e-mail servers. You need to have all these pieces together to work this fully EAI working solution for you.

In our case, we started to build anti-spam first 17 years before and then we came into e-mail. So we have our own anti-spam. I mentioned to you Spam Jadoo, which is EAI-ready. And we first converted that and see that can we have some exchange of e-mail without [disrupting] the mails and all that stuff. We first converted that and our anti-spam is two-way, which allows you to send out an e-mail and receive an e-mail. So that was very easy for us to move the traffic to anti-spam and try it out.

For us, the good part was that Google, Gmail were supporting that. So testing becomes a bit easier, so that if there's somebody who is receiving and you can just [inaudible] receive an e-mail back. But almost nobody large player receives that.

Unfortunately, my Outlook account seems to be on an older datacenter. I'm yet to get a support on that. Yahoo! obviously do not support that. [Reddit] do not supports – nobody supported that at that time and we even – we started building.

By 2016, government of India came out with very magical announcement that now you can have a domain name completely free. Don, that's the reason I sent you an e-mail out so that you can have a domain name completely free. So Mohan probably paid for it, I paid for it but you all can have a domain name free. Completely free.

We have a website called godil.in. There are [inaudible] like that probably but one website godil.in. You all can go and register a domain name for yourself completely free in [inaudible] and these Indic languages.

This also gave us an opportunity to think that why can't we start now because government is pushing this agenda that people can have it. What do they do with this domain name if you do not have at least the communicability on that domain name?

So that forced us. We were already working on that. That forced us to speed up the entire initiative and bring it something, which we can do. So we had a huge discussion on the board that can we launch it free, paid, what do we do? So the board gave us a clearance that let us do it free like Gmail does in English, let us do it in Hindi. Let us do it in [inaudible] languages.

So, Mark, just to give a small correction. We are not seven plus one, we're eight plus one. So we have 12 languages: eight Indic, three international: Russian, Chinese, and Arabic. Three and one English. English is just there because we didn't want to be – discrimination with English. That's just there, so not underplay with that. Nothing else. We are not interested you signing up in English at all. So better to sign up in another language and that's doing very well pretty much.

As Don wanted in Hindi first, he could not change his DNS, so he gave it – or do Arabic domain name, which I don't know what it is. I need to translate and see what it means in Arabic. But we can just technically configure and give it to him. I don't know your e-mail ID, but when you send me an e-mail, I can reply and that's it. So it is…

DON HOLLANDER: In Arabic, his first name Don has become Dun.

AJAY DATA: So I always have a transliteration problem, always mine – my surname is Data and when I transliterate it, it always gives me data. It never gives me Data. And actually it is Data. But now, I think Google Keyword is intelligent. He has now learned that I always needed Data, it gives me Data now. But when you do it first time on any other computer, it always gives me data. So I think these are the challenges, which we already talked.

Some of the things, which I want to bring to the knowledge is iPhone, which we are not talking. It does support EAI at all. You can't configure your EAI ID on that. We need to talk to Apple for this. I would say it's a huge thing which we need to get it done in the new initiatives so that the support configuring the EAI

mailbox on iPhone. That's the reason we had to do our e-mail app. Yes, Don?

DON HOLLANDER: Yeah. Just to be fair to Apple because there's nobody here, they are working on it.

AJAY DATA: Great. Perfect. So I'm sure they must be working. Today, it does not support. And I want to know also you're using Mac, what about the Mac Mail? Is it you're able to configure your own Arabic ID on Mac Mail or no, not right? Not right. I would like to hear that, how does Mac Mail behave.

[Rajiv] [inaudible] sitting here and [Reshmi] sitting here, I would request that if we can have it attached with UID. So anybody who has an [inaudible] card, if you can have an ID from now [inaudible] now reaching to billion people if they can have the linguistic e-mail space, the government can communicate to them very much more easier.

It's very, very simple so we can just allow people to sign up. Like [that e-mail], I am sure a lot more services will come. People have signed up for Gmail when we ask the e-mail ID, so people can sign up in Hindi ID or a Tamil ID and they can be with UID. That can be an amazing thing if we can have billion IDs, and the

government enables to communicate and allow them to [inaudible]. That way, they will also come on board with Internet.

For passport and for airline sites, another problem, which ICANN support is required on Facebook and Twitter. This is also a huge initiative which required that both do not support the sign up with EAI today. We need to have them on board because they are most used social media sites. Instagram also for that example. Pinterest also for that example. They're all need to be communicated.

And I think once we have them announced only, IDN [inaudible] awareness, which Ram was talking will automatically go and shoot to everybody's mailbox automatically. If we just [accept] these four companies and just let them announce that they are now EAI-ready. It will clear the curiosity, clear the awareness, and people will look for the regional language as the e-mail ID.

Outlook 2016 very well support. Thank you Mark, Microsoft, which is very, very neat work. It works superbly well but not [your] Outlook app. I hope this will get a very quick update, Outlook app on Android is still…

MARK SVANCAREK: I am unable to give you the dates for unreleased software, but we are working on it hard and it is on track.

AJAY DATA: Great.

MARK SVANCAREK: It is [against] the schedule that they put forth, so things are good.

AJAY DATA: Great. [On the way]. So coming back to Data Mail and XgenPlus, XgenPlus is fully compatible with unicode@unicode, punycode@punycode, complete compliant. We do not support unicode@punycode. We [told Don] in the last session because though it's the standard as I am told, but it doesn't make sense because either on an e-mail server side either I do support IDN or I do not support IDN or Unicode for example.

So on Yahoo! our tests failed and the initial when we were developing because when you deliver an e-mail, unicode@punycode, Yahoo! does not support it. It don't perceive it. But you have to have punycode@punycode to deliver an e-mail to – and there are lots of – and the government servers are like that today. You cannot deliver a punycode.

Display names are fine. Display names can be anything. We just [today] all the e-mail servers supports it. Display servers – display name is not a problem. It is about the local part of the e-mail address, which is required to be punycode. I have rather suggested that it should be a standard. Punycode@punycode makes sense, which I would like to write officially around it.

Contact address book is a challenge, a lot of – because view – everybody is [inaudible] and validations and search. So it does not search during, then you become search. And when you are Skyping, you have to be intelligent in which language you are typing.

Probably let us say your contact address book has Ram it does in Tamil and Don it does in Arabic and my address in Hindi. And you're typing in some character set. You have to be intelligent enough to understand how to search. All these technologies have to talk to each other and get the desired result.

This took a lot of time for us actually, Don. This took a lot of time to get the web interface ready for searching and giving the right e-mail addresses on your interface and to CC, BCC. [By the way] to answer your 2 comma, you can put comma in your interface and [inaudible] to anything.

On mobile app, I would encourage everyone, the entire community if you're hearing online also to me, that you can

download the Data Mail app and start testing your EAI applications. This is actually the need of the day that we can all exchange the environment. Thank you for putting up the e-mail yesterday of EAI compliant providers. That was very helpful. So we can talk to each other and fill the gaps wherever they are and bring more people online on the linguistic space. Thank you very much.

DON HOLLANDER: Thank you very much. Ashwin, how did ICANN IT approached the UA program of work.

ASHWIN RAGAN: Thank you, Don. This is a seemingly daunting task when first you approach this problem because it's a problem how to find things first and then having to fix them. For regular companies that are managed, it's not as easy as it seems only because even getting an inventory of what needs to be reviewed so that you can find things is oftentimes the first challenge.

As the company size grows, these assets can be spread in many, many different ways. If companies will progress to the point to having what is oftentimes termed as Shadow IT, then you really have no central repository where all of the IT information assets are cataloged and controlled.

So having that as the first step is a prerequisite to succeed in this program, and that's sort of where we started. It was fortuitous that the time period that I got to ICANN coincided with the trust that we have with Universal Acceptance. We'd had history of some Shadow IT, and some of our assets were indeed in different parts of the company when I got there.

The first thing that I had independently started to do was to assemble all of those into a central repository so that, with that inventory, we have the opportunity to start looking at what the extent of work ahead of us would be.

Now, what's interesting as time unfolds is that many digital services are becoming in the cloud services. These are software, the service offerings that companies subscribe to, they leverage these services.

What's also interesting is that while the service itself might be a standalone, the data models are often shared between in-house services and cloud-based services where you could have data that's coming in from the cloud be stored in a common repository from which an in-house service may be picking things up.

We found evidence of that as well. We had a number of our services actually as software as a service that we were

subscribing to. A number of them were reading and writing to in-house written applications.

Then, we looked that the in-house application suite and found that over time, as most if not all shops do, the number of platforms we'd leveraged were many and varied. We'd started out back in the day with a suite as software.

And when I talk about a platform, even if you take something as simple as Java for instance, we had multiple versions of Java in production that had pulled down multiple libraries in the art of creating different applications.

So it wasn't a case of we fix Java. It was a case of we fix Java version… and it pulled library XYZ from repository ABC. If we had code control over the library, we had to go back and correct that.

So inventory, step one, was followed by assessment. This literally was a stage of assessing what is the extent of work required and what all the variations in that assessment revealed to catalog those clearly.

Then, we started to talk about what steps could we take because trying to swallow this was like trying to eat an elephant in one sitting. You can't eat an elephant but you can eat an

elephant sandwich. So we said, “How do we start slicing pieces of this elephant and which slice makes the most sense?”

We decided that we would look at three or four more commonly used platforms that were internally leveraged to build our digital services and learn our lessons with one or two programs in each of those.

So we picked on a combination I think of four platforms finally. And between those four, we had a sample size of 10 or 12 programs that we opened up.

Now, one of the things that happens in companies like ours is that this code is in production and when you open code, it is not always going to be what you wanted to be. I mean, it does things that you don’t wanted to and the programmer gets interested in fixing things that you’re not asking them to, and bugs that have been there are tempting now to be fixed because you’ve got the code base open. So there are all kinds of considerations.

So we started to put some very tight fences around this saying, “You can do this but you cannot do that,” because our primary interest is to become UA-compliant, not to fix old bugs that have been [in there]. We don’t want you to create three more bugs because you wanted to fix the one bug that’s been there for ten years. So we have to put those time fences and boundaries around what could and could not be done.

As we went down this path, one of the things that are requested of our team, and [Sig] did a brilliant job of this, was to say, "Write everything down because we will find things that we didn't expect and we will expect things that we won't find."

So just write everything down, treat this as like we're a laboratory and it needs to be like a lab information system manual where you're writing every step down so that we don't go back and rediscover the same thing the next time we do this stuff.

And my colleague, they did a great job with that. When the first success happened, it was a moment of pure joy because it took a heck of a lot longer than we thought it would.

So I think Ram alluded to this saying but it may be a half a day's fix, but it will take a week. In our case, it was a half a day's fix and it took like a month and a half because it's like every time we poked, there was something else that needed to be touched and every time we touched it, it was leading us to some other place saying, "Poke me here because you'll find this here, too."

So we went farther and farther back in the code base until we finally got it. And when we did, it was a moment of sheer pleasure and joy. I think the team celebrated that very appropriately and then we said, "Okay, one done. Good golly. A hell of a lot to go."

So it's been a long time in the coming. We created this as a very tight program and approached the first learning with that program in mind. About a year later after we started, we were at a point where we declared victory with the program as a pilot and said, "We have the 10 programs or 12 programs that we initially had in our gunsights completely taken care of."

We then took a step back and said, "Should we continue addressing this across the entire portfolio or what do we do in order to get UA readiness?" And strategically, we took a decision that this was a) critical, b) not necessarily the only thing that we needed to do because we have competing priorities like every other company in the world, and therefore c) that as we open up code bases for any other purpose, UA readiness would become a common story that is written into every program so that no matter what reason caused us to open up a code base UA readiness would be one of the things that would be addressed so that when the code base put back in production regardless of why it was opened up, we would be UA-ready as well.

So that's our story so far, and we are making good progress with the continuing program too. Thank you.

DON HOLLANDER: Thanks very much, Ashwin. I think you raised some very good points and one of the very interesting things that we've heard

from Sig was one of the unintended byproducts of this is he has an excellent data dictionary. So he knows every data element he has, its characteristics, and where it's used. I don't know how many people here are CIOs or just had been when they were younger perhaps system architects. But this is something that you are always aiming for and life gets in the way.

So Sig how steep was your learning curve, and how quickly did it take for it to plateau or has it plateaued?

SIG: I was thinking about that and I took the question just sort of as for other people. Will I have deep learning curve, and will it always be steep? I figured my answer was yes and no and yes. We had a steep learning curve because we started a year ago and we were doing R&D. This was an R&D project to say what will it take, we didn't know. To be honest on Wednesday, I learned something we are not doing yet which is the dot in Chinese is a different character than the dot we use. I'm not doing that, so back to the drawing board a little bit.

But that steep learning curve, a lot of it is done, and I don't mean just done for us. When this group publishes lessons learned, then the next company doesn't have that learning curve. All we did was break code and find out why and fix it again. That list of lessons learned that this group can publish not just from us but

from Microsoft and from others will eliminate or severely reduced a lot of that learning curve. That's the no, that was the yes and then the no.

The last yes is though, and Ash called this and it hasn't changed yet, that year and a half ago, we talked about this and he said, "This is a Y2K problem." We go through a lot, and it's still a Y2K problem.

The code is interesting. It's not that easy, but it's not that hard. We are not getting our best programmer to figure out how to do this but finding every place it touches is amazing. And that's the data dictionary. My front end is in Ruby but my back end is in DotNet or Java, and by the way I've got a program over there in Drupal that pounds on the same database. And I converted the database from ASCII to Unicode – and I just broke somebody up.

That type of consistent use is the back to yes. That's not a learning curve but that's a path. That keeps it from – unless you have a really nice confined and [inaudible] and I appreciate when that happens – can find applications when they're not that. You do find your way tracing back. When you get that data dictionary, a lot of other things get safer and easier later.

I want to say one thing about the data dictionary: you may want to do it, but you'll never get the budget for it. Here it comes, a piece at a time.

The learning curve can be steep, but I think this group right here is going to make a real big pass for companies at making that a lot shorter. Again, just like Y2K, there is a process involved in finding all of the parts of the spider web. That's just part of what happens. At ICANN, this isn't should we do it or shouldn't we. This is we're doing it. Just how do we do this as quickly and safely as possible?

DON HOLLANDER: Thanks very much, Sig. Sarmad has told me that I'm overtime already, but I'm saying we started late. If I could have five minutes for Q&A, Sarmad if that's all right. Are there any questions?

UNIDENTIFIED MALE: I have a question.

UNIDENTIFIED MALE: Thank you, Don. This has been very interesting session. What I would like to know is what are the timelines? Mr. Ram Mohan pointed out that many jigsaw puzzles are already in place. But how we integrate it? Is there any timeframe or any kind of roadmap which we can chart out for Indic languages which will help out to track down the things other colleagues also around

the table here? So that we come to a definite time period to reach certain place. Thank you.

RAM MOHAN: [inaudible] I think there are two parts to this. One is in the UASG we have done quite a bit of work in defining what does it take for you to get UA-ready. That's the foundational work. That defines some of the technological pieces that have to be done. The part that is missing and that we need engagement from inside of India is, who do we have to get to the table so that we can move that forward?

In my mind, what we need next is a workshop here in India that is focused on – you can call it Indic Computing – but it really has to be a workshop on universal acceptance of domain names and local language content. In that, I think what we need to do is to bring developers. We have to bring content producers. We have to bring companies, webhosting companies that are here in India who provide services, and the customers. In my mind, those are the three top groups that have to be brought together.

But I think that if we can host a workshop where we walk through two parts: one is the foundations of what is an IDN, etc. Many people already know that. What is more important is, how do you make your websites and your e-mails actually accept? How do you ensure that when you store a domain name in

Unicode that you don't mangle it in your database? Those are fundamental things, but they actually are not common knowledge. That is what I think the workshop should focus on. We get that going, I think we will actually make a lot of progress.

In my mind, in a 12- to 18-month time frame, we ought to convene again and see where we stand. But that baseline is what the first workshop I think should do. I think IAMAI is a fantastic place to help host that.

UNIDENTIFIED MALE: Ram, can we do it in [Japa]?

RAM MOHAN: [Inaudible].

UNIDENTIFIED MALE: [Inaudible] a year like we can all have holidays and do some work in the day. So this is a lighter side of it, I had a question with Ashwin, and you practically told that whenever you are touching the code, you have a mandate to do UA compliance in that code. If you do not test the code then – because there will be some very important pieces you might not touch. Like for example, I had a question with Don also. I subscribed on uasg.tech with my Hindi and my English, both. So I don't know

what happened with my Hindi, but I got list [edit] of my English ID.

DON HOLLANDER: Let me answer that question and then Ashwin. So we use Mailman for our mailing management. Mailman is an open source mail management tool, been used for years, used for million. It is not UA-ready. So I do the same thing. I have an IDN e-mail address.

RAM MOHAN: I will host it for you.

DON HOLLANDER: Let me finish. I went to register for the IETF with an IDN e-mail address. It said, "No." I wrote to the young man at the IETF. I said, "This doesn't seem right that your systems don't comply with your current standards." They said, "Oh, we are so embarrassed." Anyway Mailman is actively being worked on, and there are rumors that it will be compliant at the end of this year which is very soon. Ashwin, sorry.

ASHWIN RANGAN: Thank you, Don. Two different things I wanted to address. One, is Dr. [inaudible]'s question, is there a timeline here? When we

talk with service providers, the level of awareness is fairly low. And we have had a numerous set of conversations with service providers who are India based and outside of India too. Part of this is getting to an understanding of what's going on. But I also think that there is an economic incentive. If I look back to the opening ceremony for instance where Minister Prasad was speaking, he said that the Indian export segment is now contributing more than $100 billion. But there's about $20 billion that's inside of India because the total being contributed to the GDP is $120 billion.

There is no knowing what that upper limit is until the Indic languages are fully comprehended in whatever is being done in the space. Because there is no reason to believe that there is a monopoly to English speaking people on creativity. I think increasing that level of awareness will create a natural urgency.

While there is no Y2K ticking time bomb that says on the 1st of January 2000-something is not going to happen, there is an urgency when people realize that there is potential that can possibly be looked at from an economic upliftment perspective. I think that is an angle perhaps that could be played up. The capabilities are there. There is no doubt about it. With things like what you're doing, Mr. Data, then it becomes that much faster.

The second question that I want to answer is your question about what happens to code bases that are not being opened up. In all companies, there are multiple ways through which the portfolio is reviewed. We look at things like where is the highest likelihood of an impact in addition to what is being currently opened up because there is a project.

It isn't a singular path through which we open code. It is through multiple paths that we decide to open code. But no matter when we open code, this is one of three things that we do actually now that you've asked the question. One is UA compliance, the second is IPV6 compliance, and the third is DNSSEC compliance. We do that as a matter of routine.

And because we're doing it with our own software, one of the other things that we're now doing is where we have leased software or software as a service, we write these three as contractual obligations on part of the vendor. So we force that on the vendor. Now not all suppliers are in a position to comply immediately because many of them don't even know that these things are happening. But we sit them down at the table and say, "If you don't have it, give us a timeline. Create a roadmap. Tell us when you will be done because when we come up for a contract renewal, if you are not done, then we'll look at the competition. So you have to earn our business." So that's part of the pressure that we're putting back on the industry. Thank you.

DON HOLLANDER: Sarmad is giving me evil looks.

UNIDENTIFIED MALE: I don't believe that.

DON HOLLANDER: At least in his mind, I know it. So our goal of this session was to let people know about the issue, to let you know that it is a programming issue. What we're hearing is it's complex but it's not that hard. It is not that big deal, and as has more and more people start pursuing that, then the libraries will get better and better and better. The goal here is not to talk to you about the UASG but we are writing a white paper to address the economic, social, and cultural benefits of being UA-ready.

We do have a project to review the most common computer programming languages and libraries and make those UA and the EAI-ready. So we have lots, but if you want to know more about our UA initiatives, the website is uasg.tech. We have these brochures, Sarmad, I'll leave you to the next section. Thank you very much.

SARMAD HUSSAIN: Thank you, Don. So we're now going to move on to our next section on Internationalized Domain Names. We are now taking one step back from UA and we will, in this section, try to look at how domain names are actually enabled and what it takes in different scripts across the world. We will actually look at the initiatives going on in India in that context.

In the first talk, we have Rinalia Abdul Rahim from ICANN who is going to be introducing ICANN's efforts towards promoting IDNs in local languages and scripts across the world.

We then have Akshat Joshi from C-DAC. He is representing Neo-Brahmi Generation Panel. He'll be presenting on the work they're doing here in India on developing root zone label generation rules, which will define valid top level domains and variants of those valid top level domains for languages and scripts used here.

Then we will have Rajiv Bansal from NIXI. He is going to update us on the current initiative at NIXI to enable top level domains for various languages and scripts used in India.

We will then conclude with one more presentation from Gyan Gupta who is CEO of Dainik Bhaskar Digital. I guess more from a user perspective on how these domain names, which are being made available in local languages and scripts, they eventually contribute to the local businesses and end users.

With that, let me just very quickly upload the first presentation, and I hand it over to Rinalia.

RINALIA ABDUL RAHIM: Thank you, Sarmad. Good morning, everyone. My name is Rinalia Abdul Rahim. I am from Malaysia. I am a member of the ICANN Board. Sarmad asked me to make this presentation on behalf of ICANN because the Board committee or the Board working group that I participate in, which is the IDN working group, at the Board level oversees the IDN program for ICANN. Once he sets up the slide, then I will run through it. The details of what I will be talking about will be explained by the individuals who are presenting after me.

I see Akshat is there, and I know that he will speak about the Neo-Brahmi Panel experience and the others on the root zone label generation rules aspect as well. The technical people are here to answer a technical question. I am not the technical person, okay?

Here are the slides. Let's go to the first one. There are more than 3 billion Internet users in the world right now. The largest proportion is located in our region, the Asia-Pacific region. But if you'll look within our region itself, Internet penetration is still quite low.

Nonetheless, we expect that the next billion people coming online, the majority would be coming from our region. We see a lot of interest, demand, and dynamism in this region that we are in right now. Next.

Domain names, IDNs are essentially domain names which can be represented in different scripts such as Arabic, Chinese, Devanagari, and Thai. I want to digress a little bit by saying that there are more than 7,000 living languages in the world today, and 50% of them are spoken in our region in the Asia-Pacific region.

And there are numerous writing system or scripts that support those languages, but the number of the scripts are fewer than the number of languages because some languages share one script. As an example, the Arabic script is used by more than 50 languages across the regions of Africa, Middle East, and Asia.

We have IDNs, and IDNs represent a serious innovation for the Internet because they enable opportunities for local communities to come online. You can use your own script, basically, to use it in e-mail and to use it as a resource locator in web addresses. I think that helps local communities tremendously, particularly when they are using local keyboards. Let's go to the next slide.

This is an overview of the IDN program at ICANN. There are essentially two categories: IDNs at the top level and IDNs at the second level for generic top level domains. Let's focus at the top level first. Under the IDN TLD program, we have the label generation rules, which I will touch on next. There is the label generation rules toolset, which has been worked on. I think the IETF has been working on this as well. Sarmad, is that right?

SARMAD HUSSAIN: Standard.

RINALIA ABDUL RAHIM: It's a standard that has been developed. Then there is the IDN variant TLD implementation, which we are working on currently and which derives from the work on the LGR. Then there is the IDN ccTLD Fast Track process implementation. You will see the results of that in a subsequent slide.

For IDNs at the second level for generic top level domains, there is work on IDN implementation guidelines because otherwise it would be the Wild, Wild West on the Internet and we would want to see some standardization for the benefit of end users.

Then there's a reference second level LGRs as well. I think this is also an attempt to, basically, make things more uniform even though there is a lot of diversity in the Internet. Then there is

community outreach and involvement which is essential for the LGR program because it's focused on scripts which are used by language communities. Next.

Here's the result that I was talking about from the country code top level domain, fast track process. At the moment, we have 57 ccTLDs for 39 territories and countries. If you can see the text on the slide, you can see that they come in various scripts. That's quite good, but I think maybe we have more to go. Next.

This is just an overview to show you what differences between an ASCII label in Latin script where you see www.cafe.com and an IDN on the right. In what script is that, Sarmad?

SARMAD HUSSAIN: Hindi.bharat.

RINALIA ABDUL RAHIM: Hindi.bharat, excellent. I think you have an understanding of what this is and what it could be in different scripts as well. Next.

Here's the IDN TLD program on label generation rules for the root zone. This program cannot run without having script communities come together and to decide or to determine what would be valid and unique IDN top level domains. What the work requires is that you need a group to come together, you need to

have the right set of experts in it, and you have to go through the entire repertoire of the alphabets, letters, and characters in your writing system.

for each one of those, there's a code point attributed to it. What you have to do is together you have to figure out which code points would be valid for use in the root zone. Are any of those code points variants of each other, meaning that they are interchangeable in any context? You have to figure that out. Are there any additional constraints under label itself? Next.

This is the status of the script community work that ICANN is facilitating at the moment. There are more than 28 scripts that are being worked on with more than 19 generation panels.

The generation panels are essentially the way that we call the groups working together. For example, I was a member of the Arabic generation panel that worked on the Arabic script that involves more than 50 languages. And the Arabic generation panel was the first one that was launched and set the template for this entire program. Next.

In terms of the root zone label generation procedure, we could have several generation panels working on the different scripts. We have a panel of experts called the Integration Panel. They will look at all of the outcomes or the output of the generation panels and look at it and see whether it complies with the rules

and procedures that have been set in place. Then those that comply will be compiled, integrated into rules for the root zone. Presenters after me will go into those processes in more detail.

There are some challenges related to working on scripts. There are some writing systems that share the same scripts, so you need to coordinate because there may be variants and it could lead to confusing end users. As an example, the Chinese, Japanese, and Korean languages all used the Han script. They need to coordinate in terms of what do we do about these shared characters.

Then some writing systems may have cross-script variant letters, and you need to coordinate to identify them and it can be quite complex. As an example, within the generation panel – the Neo-Brahmi one which Akshat is working on – you have all these other scripts that you have to grapple with Bengali, Devanagari, Gujarati, Gurumukhi, Kannada, Malayalam, Oriya, Tamil, and Telugu. Then in other example, across generation panels where you're actually working on different scripts but there are some similarities: Armenian, Cyrillic, Greek, and Latin. Some letters look the same in other scripts in this set.

The challenge is if you can't sort this out, it cannot go into the root zone. Even if your script community has worked on it and you've finished, but if there are coordination issues, you cannot

really proceed. Nonetheless, the work is valuable. Let's go to the next slide.

This is a slide on how to get involved in the work. ICANN is always looking for volunteers. You can volunteer for the generation panel of the script that you are using.

As for myself, my country Malaysia, their official language is not Arabic. But the script Arabic is used to write our national language which is also using Roman alphabet, Latin script to write. And that's why I got involved focusing on the use of Arabic script in the Southeast Asia region.

There are other people who participated based on their interest as well. It's not for personal interest, but it's for the benefit of the community in your region or in your country.

Then you can do review work through public comments. If you don't have time to participate individually in the generation panel itself, you can wait for the work to finish its different milestones. ICANN will always put it out for public comment. Take a little bit of time, have a look at it, see if you think there are challenges there that pertain to your own language and script and you can flag that to the ICANN team. They can take care of that moving forward.

You could also just keep yourself updated if you are interested in this topic. I think that's the last slide, right, Sarmad? Thank you.

SARMAD HUSSAIN: Thank you, Rinalia. We will move on to the next presentation. We have Akshat Joshi who is going to tell us the work being done in India for languages and scripts which are used in India. Over to Akshat.

AKSHAT JOSHI: Thank you, Sarmad, and thank you, Rinalia, for laying out the background for what I am going to speak about. Once we have slides on, I'll get into the details. I'm basically from C-DAC. Just before I begin, I'll give you background of my organization. I'm from C-DAC and we have been working with Ministry of Communications & IT and NIXI for the [.bharat] domain names enablement which have been launched. We are seeing that there are efforts being made by the community to increase the uptake of that, and that is really an encouraging picture.

Moving onto the Neo-Brahmi generation panel activity, can we have the next slide? When we started thinking about which way we should be going, most of the other panelist, if you see they're mostly single script panels. Whereas, in our Neo-Brahmi generation panel, we are almost tackling nine different scripts.

The reason behind that was we all know that they share a unique syntactical structure, and we want there the root zone also. There should be a unified treatment given to all of them in terms of what security considerations we take into consideration while framing the structure and variance issues related to that. We took all of them into under one umbrella and then formed Neo-Brahmi Generation Panel. These scripts are used by multiple languages on most 22 official languages that are spoken in India, and even other languages are being spoken in the world.

There are two major language families that come out of Brahmi that are Indo-Aryan that is the region in the northern side of India majorly and then the Dravidian family of languages which are spoken in this part of the world which is south of India.

Like I said already, why we decided to treat that as a Brahmi, because of the unified structure. There was not an earlier standard existing with government of India, Bureau of Indian Standard that is ISCII, Indian Script Core Information Interchanged, which had a very well-defined syllabic structure. We just took forward this research into their domain names which also facilitated variant identification for us. Next slide please.

Why Neo-Brahmi? We already have seen why we did it for Brahmi, but then why Neo-Brahmi? Because not all the languages or scripts that belong to Brahmi family are living languages. For this, we took help of EGIDS scale and only those are belonging to one to four we have taken as the languages that we will be considering. This is what we have come to.

Since the number of scripts is lesser 9 but the languages that are being spoken are large. So we took an approach that we should get maximum representation from various language experts. Also, we should have our policy experts if possible from various languages, but that was not a mandatory thing. From the way we structure it, the C-DAC team which has been already involve in this, we became core coordinating team under Neo-Brahmi generation panel, and then we have different sub panels under that. The slides are already available, so I'm not going to individual listed here. Next slide please.

What exactly comes under Neo-Brahmi? What are the cases that we are dealing with? This is the classical thing that we always talk about. We have one language being returned in multiple scripts, and we have one script and multiple languages using it. These are the kind of challenges that we are dealing with.

This is the structure of the Neo-Brahmi generation panel. We have a Chair. Under that, we have language neutral policy

experts, Unicode/IDN experts, registry/registrar operation experts. Below them, we have individual subgroups which are more language specific.

Linguistically what is the structure? Under Neo-Brahmi generation panel, we have sub panels for each of the scripts. Within each of the scripts, like for example the Devanagari sub panel, since there are almost 8 to 10 different languages that are being spoken, we have onboard the experts for those languages as well. Similar is the case for Bengali script, which is used by three different languages.

When we began, the mandate was quite large and though there was an initial study that was done under [inaudible] policy, this was a universal effort and everybody was needed to be onboard. Our initial time wanting to getting maximum number of experts onboard.

I must say it was a bit difficult to get people onboard because we were seeking the experts in the field of linguistics. We wanted them to work for this project voluntarily. As well as it was a fairly complex process to explicitly make them understand what this is all about so that their appropriate feedback comes to us. Just one slide back please.

We have been participating in various ICANN workshop conferences as well as we had participated in AprIGF for that.

Even within C-DAC, when various seminars are conducted, we try to mention Neo-Brahmi and get as much support as possible.

What exactly is the root LGR procedure, and what are the major components of that? I would now speed up the presentation because from here it starts technical, and I understand there are two sessions after this. There are code point repertoires for each of the scripts that we have to finalize. Then there are variant rules which we'll have to finalize. Mainly, the variant rules that we've seen under Neo-Brahmi scripts are visually similar characters, but we will go into the depths of it and also try to analyze the cross script variance issues which was not the inclusion in [inaudible] policy.

Then one of the major [chunks] that we have to do is the Whole Label Evaluation rules for each of the scripts. Previous similar work was the work that we did for .in registry on the framework which is behind the current .bharat domain names. Next slide please.

What were the major challenges initially for the national policy? We had a pretty narrow zone to look for and we could take liberty in framing rules however complex we wanted because we were only going to implement them.

Whereas, in the [inaudible] procedure, there were certain inherent principles that we had to adhere to. That made the

game a bit difficult. However, we always maintain that the crux of everything was the actual formation or the syntactical structures that are in the languages have. We should always be able to maintain that that was one of the major focus areas. These are just some of the glimpse and some of what is currently happening. We have code-point repertoires.

In this, ICANN itself, we have seen that there would be considerable progress. The speed of the progress would be greater because of the face-to-face interactions with the integration panel. We see the next phase of the work to be speeded up a lot.

These are some of the cases of cross script variance. Here we are seeing some of the instances. I don't have a pointer, but on the below right corner, you can see there are four big black letters. Even though at this size they don't look exactly similar, but believe me when you go to the point size of a browser and when we think in terms of a layman accessing a domain name, these things pose a challenge in terms of visual similarity. Next slide please.

We'll go into the details of Whole Label Evaluation rules. Next slide please. It is imperative that when we start looking into this, we should look at where the exact policy elements of the

previous work. I'll touch upon them briefly in the next slides. Next one. We can skip this.

These were the basic components that we had identified for Indian languages. These are the categories that we had created for strictly forming the rules. When we finalize the root LGR procedure, initially it was drafted. There was a feeling that we should be only modeling the categories after the existing Unicode categories. If we considered that as a given, the task was a bit difficult. Next slide. These are the categories. On the next slide we'll see the formalism.

These are the basic variables. This is a bit technical stuff, so I'll try to get it faster. These are the variables that we used. This is the formalism. I'm not going to the details of it, but this is the formalism which has been made as generic as possible. This is the structure which caters to almost all the Indian languages.

If you can see, there is a big line, it caters to almost of them the Devanagari based languages. Below there is Z which caters to Bangla exceptions. Below there, L is for Malayalam Chillu characters. Below that [inaudible] the case which caters to Punjabi characters. This one structure caters to all of the scripts.

The fundamental difference between the work that we had already done and what we had to undertake under the Root Zone LGR is .bharat was quite specific zone. Root Zone is the

most generic zone possible on the Internet. There, we are only restricted to Indian languages. Here, we cannot restrict to the languages, if there come the cases we may have to look at the languages which are spoken outside India.

Policies could be restricting the previous work here, policy is had to be simple yet sufficiently tight to restrict the possible spoofing and phishing cases. We could define our own categories. Here, we had to rely on the Unicode properties which has been. Integration panel actually did a preemptive strike on me yesterday, and this point is no longer valid one. But since we couldn't change the slide, that has not been taken out.

This is just a scenario. This is where the left hand side, the categories which were at our disposal. On the right hand are the categories that we thought we had it at our disposal and that made the task a bit complicated. This is where we are actually. Again, touch upon the schedule and other part which we saw under the update session. Thank you.

SARMAD HUSSAIN: Thank you, Akshat. We are now going to move on to our next presentation by Rajiv Bansal from NIXI. He is going to talk about the IDN ccTLDs for Indian languages status and vision. I'll upload the presentation.

RAJIV BANSAL: Thank you, Chair. I'm Rajiv Bansal. I work for the government of India, and I am also the GAC representative from India. Let me begin by first placing on record my sincere thanks to ICANN for having organized this session. And also to IAMAI for having brought us all together in this room. It's indeed a pleasure, and I'm happy to share with you some thoughts.

It's a pleasure to note that today in this room, we have all the critical stakeholders. We have representatives from the government, myself and my colleagues here. We have representatives from the registry. I also represent the Indian Registry. We have representatives from the industry both international industry as well as the domestic industry. We have representation from the media, and we have representation from the civil society.

A lot has been said in the last two hours, a lot of technical issues have been flagged. While I do have a presentation which focuses on some of the technical issues, I thought at this stage of the discussion it would be useful to bring in some policy perspective.

Thereafter, maybe I can speak for about five minutes on the policy issues and for couple of minutes on what NIXI is doing. NIXI is the National Internet Exchange of India. Let me begin by

saying that to my mind, the most critical driving force in bringing about this whole – in addressing this issue of IDNs is the government's desire to bring in a digital transformation.

While technical issues, technical aspects, technical solutions can be done, the single most important driving force to my mind has to be the intent or the desire of the government of the day to project a need for it, to bring in a need for digital transformation.

That is what I think the government of the day in our country is trying to do. We have a huge challenge before us. I say it is huge and I emphasize the word huge because of two things. One, we have a population of which about 30-35% we have been able to penetrate the Internet. So about 35,400 million people are connected to the Internet. This population is largely living in the urban or semi-urban areas and they speak, read, write, understand English.

The next 35% is living in semi-rural areas, and they are literate but they are literate in their own language. Yet and thereafter, we have another 30-35% population which is neither literate, not literate in their own language and definitely not digitally literate.

Until now, I have talked of percentages. The second big challenge is that each of these 35% is the size of one America. So we are looking at 300-400 million people in each of these

categories. As is often said, there are so many Indias within India. It is the second and third Indias that we are trying to address in this challenge.

We have often said that – we had a meeting with the ICANN Board also – that our interest and our stake with the ICANN process is so high today. Because it is only with the support of the ICANN that we will be able to achieve this objective. Often we find in plenary sessions it is said that the next billion users will come from India and China. I do believe that this is not going to happen unless we are able to address the issues that we are addressing today in this room. It's extremely important and I think a lot of focus needs to be given to this.

The good thing is that the government at least in India is focusing a lot in this. I would dwell on four or five aspects of the government is doing, and then I'll focus in two or three aspects that the registry is trying to do.

First and foremost for us in the country is to give access because if we don't give access itself, whether it's an English or Tamil or Indic language, it doesn't make any difference. The first and foremost thing is to give access.

It's really mind boggling, it's really challenging to see in our country that today we are laying not only the optic fiber. We are laying the electricity line. We are laying the water line. We are

laying the optic fiber. It's not only digital transformation. It's kind of a social transformation. We are making available to the people – to the marginalized people, to the rural people, to the economically deprived people – water, electricity, as well as the optic fiber at one go. To that extent, it's like a transformation. It's leapfrogging many stages that we are trying to do in one go.

We launched the .bharat in 1994. We haven't had traction on it as we would have liked, but at least we've launched it so we have taken the first step. How do we build traction now? That's the challenge that we have at hand. What is happening is that over the last decade, we had what was called the mobile revolution. A period of 6, 7, 8 years we added 100 million mobile subscribers year on year. Today, we have a thousand million mobile subscribers.

These next 3-4 years, we plan to add 100 million Internet subscribers every year. Not do it one year, but do it year on year for the next 3-4 years. How do we do that? We need to do two things. We understand that the preferred device for people who would come onto the Internet would be the mobile. That is why IAMAI, the Internet and Mobile Association of India, is a key player in this. How do we bring more smartphones to people? How do we bring down the cost of these smartphones? How do we make these smartphones more robust? That's a challenge that we have at hand.

A recent decision by the government I would like to just flag at this point. We have built a new standard which says that all mobiles sold in India with effect from 1st of July next year, 2017, would compulsorily by way of legislation, by way of a standard, have to have text inputting capability in three languages and text readability capability in 22 languages. We believe that this would help a lot because mobile being the preferred choice for accessing the Internet and there being a standard in place to ensure that all mobiles have the technical capability. We do believe that this would help in Internet penetration along with access of course. We have also an ambitious program to lay optic fiber to 250,000 villages in two and a half years.

Side by side, we have launched a program of digital literacy wherein we plan to make 60 million households digitally literate over the next few years. A lot of this that we want to do is going to happen in the years 2017 and ’18, so we have about 24-26 months left to do all this.

Another big area that we wish to work on is we have a unique system that we have been able to do is to develop a digital identity for almost everybody in our country. So today we have an identification system called Aadhar which provides a biometric database for more than 1,000 million people in our country.

We plan to dovetail this with the mobile and be able to launch all government services over the Internet and make all services what we call faceless, paperless, and cashless. We believe that with this coming together of the smartphone, of the Internet, and of the unique identification system, we would be able to leapfrog many steps that other countries have moved over a period of time.

For example, a very large number of people who were outside the banking system today have bank accounts. But we are trying that these people can transact with these bank accounts on the mobile using the digital identification, the biometric identification system that they have.

In a way, we would be leapfrogging three or four generations of banking, three or four steps that the whole world went through. People went through with post-office bank accounts and they went through the passbooks and they went through checkbooks and they went to debit cards and they went to credit cards, plastic money, and then into e-banking.

What we are trying to do is a transformation where our people in the villages take the fifth step as the first step. To my mind, it's a huge thing that we are trying to do. I don't know whether we will be able to do it and in what timeframe, but we need the support of the ICANN to be able to do this. We need the support of all the

stakeholders who are in this room to be able to do this in the timeframe that we want to achieve this.

Having said that I will now come to one more aspect. I also wear the hat of the .in registry. I am the Chief Executive Officer of NIXI, the National Internet Exchange of India. We are trying to dovetail our efforts with what the government is doing.

Two initiatives that I'd like to focus on. One is that we are providing as of now a 100,000 .bharat IDNs free of charge, gratis. Now this is something that Mr. Ajay Data also referred to. For some period of time, we're trying to get traction, we're trying to get mobility on this front by providing .bharat in the languages in which it is available free of charge to the extent of 100,000 numbers.

The second thing that we're trying to do is we have these information kiosks called Citizen Service Centers which have a huge reach. We have 600,000 villages in our country. These information kiosks are there in more than a 100,000 villages.

We have made these 100,000 Citizen Service Centers, the information kiosks, as our registrar for the .in. And we do believe that these information kiosk who are going to implement the digital literacy program will bring on people, make them digitally literate, and also give them a .in domain name or .bharat domain in the language of their choice.

This would provide the last mile, the cutting edge connectivity to the person who is living back of the beyond, who is economically deprived, who's living in a hilly area, who is living in a remote area. Our reach is going to happen through these Citizen Service Centers and the information kiosks that we have.

These are the few thoughts that I thought I would share with you. As I said again in my opening remarks, I'm really thankful that ICANN is supporting us in this initiative. I'm thankful to the industry, the international industries. Some of the big companies are represented here. I wish they would do more for their commercial interests as well as societal interests. I'm thankful to the Indian companies who are taking this interest and coming forward in assisting the government in our endeavor. Thank you.

SARMAD HUSSAIN: Thank you. Let's move on to our last presenter for this session, Gyan Gupta. He is going to present a perspective of a user of these localized domain names.

GYAN GUPTA: Thanks. Good afternoon, everyone. Coming from a publisher's perspective, I had [inaudible] which is the largest news site in the country and in recent years have become the second largest

site, which is just showing the growth of the local languages that is probably happening in the country.

We have been working with IAMAI and NIXI over the last couple of years. I still remember three-four years back, we were just trying to put the Unicode structure in place. I think from that perspective, we have come a long way forward.

But when I look today right now, from an end objective, I think what we really care is the discoverability of the content or the searchability of the content. That is what is really an end outcome of all the discussions probably that we are having. Once we have that in place, the rest of the efforts actually come together to make sure that we do that. Because if we will not also look at the entire searchability and discoverability of the content which is lots of milestone but very critical one, it may not put all the efforts into the right perspective.

One of the biggest reason I think it's taking a lot more effort is also to make sure that how do we bring all the structures together. While I understand that there are lots of challenges in the intra operability of the different scripts, and how do we make sure that the databases talk to each other, and stuff like that, [inaudible] I have only couple of data points to show and I think Rajiv did mention the growth that is probably happening in the country.

But if you want to put India into a perspective, India with the 1.3 billion population, English is about 11% penetration in the country. And Internet currently sitting at about 34% penetration. We have over 350 million users probably that are there, which means that I already have 25% of the people who are non-English users on the Internet.

If I were to put into the perspective of the next couple of years, we will add like Rajiv also mentioned, about 100 million users year-on-year at least for the next three or four years. Most of these users will come from local languages.

When I look at the local languages, Hindi is about 45% penetration, which is about we have about 400 million people in India who speak Hindi. The rest of the languages, eight or nine other main languages who have a user-base of about 60 to 90 million users.

As and when these users come onboard, they will start looking for content. They will start searching for information, and that point of time, all of these things may become hugely relevant. How do they discover this? How do they search for the information? In addition to this challenge that we might have generating content in local languages.

Because I don't think there's enough effort also being created. We're doing it, but enough effort being made to make sure

there's enough information available. Because at the end of the day, the Internet has to provide me a knowledge and a utility if you expect me to pay 100-200 rupees on my data packet. It has to have value beyond entertainment.

I think when we put a lot of these issues into perspective, how do we make our effort to streamline and to probably expedite that? Because we're already sitting on a huge opportunity area, if I may say so, which will actually extrapolate the impact and the benefit that it may have.

Whether it can be the Google trying to improve the local language source – because I think they still have long way to go out there – or Microsoft trying to make sure that all of the local languages' e-mails are being acceptable or Apple trying to make sure that Indic languages are accepted. But at the end of the day, all these technology companies and the device manufacturers have a huge role to play in the discoverability and searchability of the content.

While the good initiatives can be to make sure that the handset providers are bounded by a policy or legislature, which I think is a fantastic effort taken to make sure that, that I think will boost a lot of things. We also have to probably look in terms of other than this, how can we have other maybe brainstorming session? Where we have all the companies which may have a commercial

interest, how do they extrapolate the searchability and the reachability of the content? Because I think that will be one of the key benefits and impact of whatever we are doing, we'll actually look at it.

I'm sorry. My flight was late, so I'm kind of pushing in towards the end. But this is probably how we look at it or how we look at the landscape. First opportunity is huge, but I think we still have a lot of roadblocks to cross before we can actually capitalize on this entire Internet endeavor. Thank you.

SARMAD HUSSAIN: Thank you. Let's open the floor for questions you may have for the speakers here. Any questions? Yeah, Don?

DON HOLLANDER: Rajiv, thanks very much for that. I think that was interesting. I didn't understand it, but others did. But this I understood very much, the digital transformation. I think that's a fantastic government initiative. The government departments, in terms of their websites and their engagement with their constituency and citizens, are those currently in all the official languages?

RAJIV BANSAL: Substantially, no.

DON HOLLANDER: So there's an opportunity too? And associated with that, have they each registered the appropriate domain name for their government departments in the respective languages?

RAJIV BANSAL: Some yes and some no. That's the challenge. In this state where we are sitting today and also in neighboring [inaudible], they are very strong in e-governance. A lot of their e-governance initiatives are in the local language.

However, when you actually – for example, I'm applying for a birth certificate for a child who was born in my family. When I start typing and then keying in the name, the system doesn't accept it in the local language. The service is there in the local language, but when you're actually key in something, you need to again put it in the English and the Latin script. That's why the answer is not really yes, not entirely no also. It's a mixed bag.

For example, now there is e-commerce sites. There's a huge reach, so e-commerce is booming in the country. This reach is remarkable. You don't need to be really very highly digitally literate to buy a product on an e-commerce site. But then at the end of the day when you add to cart, you need an e-mail ID and that you cannot do in the language that you know. If you're

digitally literate in a language which is not English, you cannot make that e-mail ID in that language. So you cannot complete the transaction.

What we want is that you should be able to make that e-mail ID in Telugu or whatever language, make the password in that language, and then be able to transact that money with that biometric identification system that I have so that it's end-to-end in the local language.

SARMAD HUSSAIN: I have Gyan and then Mr. Data.

GYAN GUPTA: Yeah, I'm just adding a point to what Don's question where Rajiv mentioned. One of the other challenges also is in the inputting of forms or in [inaudible] in Indian local languages. While we might make a lot of effort to have a lot of these on local languages, but the inputting mechanism is still very difficult in local languages. That makes a lot of the effort also less impactful because if you really try to fill up a form in Hindi or [inaudible] online, it's actually very, very tedious because of the local language support done by the keyboard or the manufacturers.

That's another piece that has to be solved as then we make all lot of these forms and stuff all in local languages.

UNIDENTIFIED MALE: We might have just solved that problem. We are releasing a keyboard which you can embed within your digital assets which allows people to select the language and type and write there. That's going out free.

UNIDENTIFIED MALE: Actually there are two kinds of audience. All the time, we will deal with the people who knows English and other language and who knows only other language which we are talking about. Those people who do not know English have their digital assets already in that language. Whether that is iPhone in Russian or iPhone in Hindi, it's completely in Hindi and you can input. You don't need actually any other tool. For the English people who want to work in their languages, you will need these kind of tools or Google Indic tool or whatever tools.

And we are seeing both kinds of audiences here in our e-mail thing. Lot of people who have their entire phone in Hindi do not come back with any kind of issue. The issue comes with the people who are actually not my customer but come back with an issue that I want to have a Hindi ID for trial, and I do not know how to type in Hindi.

But actually we are not solving the problem for him. He already knows how to communicate. We are solving a problem, we are all trying to solve a problem for those people who do not know English. The problem for English is already solved.

So only one suggestion I can have for Rajiv is that we can at least have a domain name like India.bharat which points to [India.gov.in] so at least the IDN gets promoted. Forget about everything else. You already have a Hindi in [India.gov.in] that is already in Hindi, but we do not have Hindi domain name. Maybe we start with that. We obviously know nothing can be solved from zero to hundred in one go, but at least we will have some start. All the government sites can have the IDNs to resolve.

If people are interested to who are knowing only Hindi and [inaudible] can say [inaudible] India.bharat is not available, people can type in their mobile, India.bharat who do not know [India.gov.in] at all to type. This will at least make a lot of change and bring them online. It is a very simple thing to do. You do not need anything to be done there. Simple resolution to be done the same site. I think this is a big change which will all bring if the government start doing this small thing for all the websites which are available and yourself as a doer.

E-mail us of course very, very critical. E-mails are going to be available Microsoft has announced very soon. We are already

there, and people can have it free. This is likely to be free as e-mail businesses, so e-mail is going to be there. I am told that other departments are also working. But IDN adoption is very, very critical for all of the companies. Like when I say datamail.in, datamail.bharat in Gujarati, Punjabi, Tamil, Telugu, all resolves to in their own respective language.

That's very, very important and we see not very high traffic. Obviously, not everybody is adopting, but people access them through that by directly typing and datamail.bharat in Hindi.

UNIDENTIFIED FEMALE: Sir, I have a question for you. A lot of things are being done by the government to promote the IDNs in even local languages, even the business is also working on it. What would you look at as the top three things which you would possibly want a business also to work on, to make the things which, especially the local languages, the IDNs or the domain names in local languages work?

And also another subsequent question like C-DAC is doing a lot of work on this particular field. Is the business or can the business also be included? Because many times they are not aware of what's going on. Can there be some collaboration? Perhaps it could help us, the [countrymen].

AKSHAT JOSHI: First is that we have been working with the browser community, with the search engine community. We've tried to impress upon them the need to bring this and modify whatever standards have to be done and make available e-mail IDs in Indic languages. That's the first and foremost.

We're asking mobile manufacturers to make the smartphone cheaper. If the smartphone becomes cheaper, access increases – 70% of people in India do not own a smartphone. They own a feature phone. If the smartphone becomes cheaper, it becomes more robust so it's kind of if it falls, it doesn't break kind of thing. That's something that will help this all ecosystem.

The whole aspect of digital literacy, so a lot of big companies have CSR activities in which they do digital literacy. Our numbers are so big, everybody says that we have been doing this and I thank them for what they are doing, but it's still too much of a drop in the ocean. If we can actually make big impact on digital literacy, that would go a long way.

I keep telling all these big companies because our markets are much open than let's say the Chinese markets, we have allowed all the big American companies to have a presence in India. It just dovetails with their commercial interests to have Internet penetration.

To sum up the three points, I would say all the e-mail providers and the browser community, if they can modify the required standards so that all these Indic languages, the IDNs work. The mobile manufacturers to come forward because the numbers are big if the mobiles can become cheaper, the smartphones can become cheaper. And the third is focusing on digital literacy. These are the three big things, and it will all work.

SARMAD HUSSAIN: Thank you. I'm sorry, we'll have to close this session because we are running out of time. But before we do, I'd like to thank on behalf of ICANN to IAMAI for collaborating to organize this session. And let me hand it over to [inaudible] to make the closing remarks on behalf of IAMAI.

UNIDENTIFIED FEMALE: Hi. If I start making closing remarks, I am afraid we'll go right through to the next ICANN meeting. The list of [inaudible] is very, very long, and I'm very happy about that.

We will work together with ICANN. First of all, I'd like to thank Sarmad and it was he who was behind this. He is the mind behind this session. It came from him, so thank you so much, Sarmad. This was much required.

Also look forward to continued inputs from Sarmad and Don, and of course the rest of the industry. Mr. Rajiv Bansal, I am afraid you'll be seeing a lot more of us. We would be conducting these workshops and if Dr. Data has [his way we will] probably start off with [inaudible]. That's a promise.

Thank you all for coming and look forward to working together.

SARMAD HUSSAIN: Thank you.

[END OF TRANSCRIPTION]