ICANN85 Mumbai | Forum de la communauté – Comment ça marche : Noms de domaine internationalisés
Mardi 10 mars 2026 – 09h00 à 10h00 IST

CHAMPIKA WIJAYATUNGA Bonjour à toutes et à tous. Champika au micro. Nous allons commencer cette séance. Veuillez lancer l'enregistrement, s'il vous plaît. L'enregistrement est lancé. Bien.

Bonjour à toutes et à tous. Bienvenue à cette séance sur la façon dont les noms de domaine internationalisés fonctionnent. Je m'appelle Champika et je suis le responsable de la participation à distance pour cette séance.

Veuillez noter que cette séance est enregistrée et qu'elle est régie par les normes -- par le code de conduite des participants de la communauté de l'ICANN, les normes de comportement attendues de l'ICANN, ainsi que la politique antiharcèlement de la communauté de l'ICANN.

Veuillez, s'il vous plaît, respecter les règles suivantes. Si vous voulez participer à cette séance, je vais également les poster dans le chat Zoom. Lors de cette séance, les questions ne seront lues à voix haute que si elles sont soumises dans la fenêtre questions-réponses sur Zoom. L'interprétation lors de cette séance sera garantie en français, anglais et en hindi. Si vous souhaitez vous exprimer, lors de cette séance, veuillez lever la main. Les participants sur site utiliseront un micro physique. Toutes les

Remarque : Le présent document est le résultat de la transcription d'un fichier audio à un fichier de texte. Dans son ensemble, la transcription est fidèle au fichier audio. Toutefois, dans certains cas il est possible qu'elle soit incomplète ou qu'il y ait des inexactitudes dues à la qualité du fichier audio, parfois inaudible ; il faut noter également que des corrections grammaticales y ont été incorporées pour améliorer la qualité du texte ainsi que pour faciliter sa compréhension. Cette transcription doit être considérée comme un supplément du fichier mais pas comme registre faisant autorité.

questions soumises dans la fenêtre questions-réponses seront lues à voix haute lors de cette séance, si le temps le permet. Veuillez donner votre nom pour l'enregistrement, ainsi que la langue dans laquelle vous allez parler si vous parlez une autre langue que l'anglais. Veuillez parler clairement et à un rythme raisonnable.

Je vais maintenant donner la parole à Pitinan.

PITINAN KOOARMORNPATANA Merci. Merci Champika, et bienvenue à toutes et à tous à cette séance.

Donc, je m'appelle Pitinan Kooarmornpatana, du programme sur les IDN de l'ICANN.

Aujourd'hui, nous allons partager quelques informations avec vous. Tout d'abord, nous allons passer en revue ce que sont les IDN, les noms de domaine internationalisés. Ensuite, nous allons passer dans les coulisses et parler de la façon dont on documente les règles techniques concernant les IDN. Ensuite, nous allons inviter les membres de la communauté du Canada à nous informer concernant leur travail sur les langues des Premières Nations au Canada. Ensuite, nous aurons d'autres personnes qui interviendront pour nous informer des progrès de leurs travaux. Donc, notre programme est assez chargé aussi. N'hésitez pas à poser les questions, mais peut-être que nous n'y répondrons qu'à la fin de la séance.

Pour commencer, donc, nous allons d'abord vous parler du processus pour l'utilisation des IDN pour les langues et les différents scripts.

À quoi ressemblent les IDN ? Alors, vous voyez ici en haut un exemple de nom de domaine internationalisé, donc un IDN, pour le parathpasha.parath ; le voilà, c'est le nom de domaine à l'écran.

Donc, vous voyez ici en haut de l'écran, la forme de ce nom de domaine avec la forme Unicode, donc, ou l'étiquette U. Et en bas, vous avez l'étiquette A. Donc, c'est-à-dire que le nom de domaine du dessus est traduit en ASCII, donc en code standard américain pour l'échange d'informations.

Alors ici, vous avez différentes normes concernant les IDN, les noms de domaine internationalisés, dans les applications qui datent de 2008, qui ont été développées par l'IETF. Et vous voyez ici à l'écran le numéro des protocoles qui concernent ces IDN. Donc, pour la définition, les documents cadres, le protocole, les points de code Unicode, etc.

Alors, quelques exemples. Donc, les chiffres, par exemple, ici, donc le RFC 5892 concerne les propriétés des codes point Unicode, en fonction de la catégorie.

Donc, il y a quatre catégories de lettres. Tout d'abord, il y a le protocole valide ou ce qu'on appelle PVALID, que l'on peut donc utiliser dans le DNS. Par exemple, ici, vous avez un avec un accent aigu. En code Unicode, c'est U plus 00E9.

Ensuite, deuxième catégorie. Une règle contextuelle est nécessaire. Ça veut dire qu'on peut utiliser ce caractère, mais qu'il y a des conditions afférentes. Par exemple, dans -- bon, peut-être qu'il est un peu difficile de voir la couleur rouge ici à l'écran, mais il y a un point qui, en code Unicode, est 00B7 et on ne peut l'utiliser que s'il est utilisé entre deux L. L'idée ici est d'éviter toute confusion entre les domaines de premier niveau et de second niveau.

Il y a également une troisième catégorie non autorisée. Par exemple, le point d'interrogation, ici, que vous voyez à l'écran, parce que parfois on se demande si cela représente l'étiquette ou tout simplement une question de syntaxe.

Et puis, ensuite, vous avez une quatrième catégorie non assignée, non allouée. Donc ça, ça veut dire en fait qu'on prévoit des points de codes qui pourraient ne pas être utilisés, mais qui sont prévus dans les tableaux au cas où.

Alors, par rapport à cette norme, il est dit que, si le caractère est PVALID, ça ne veut pas dire que l'étiquette peut être directement utilisée dans le DNS, parce qu'il pourrait y avoir des règles à prendre en considération, outre la valeur PVALID de ce caractère, de ce point de code. Raison pour laquelle nous devons élaborer des règles de génération d'étiquette, les LGR. On va vous en parler tout de suite.

Alors, les LGR, les règles de génération d'étiquette, pour vous dire ce que c'est, eh bien, sont les règles qui nous permettent de former une étiquette en un script ou une langue bien particulière.

Il y a quatre étapes à ces LGR. Tout d'abord, on a les métadonnées, c'est-à-dire l'explication de ces règles. Donc, la version, la date, la langue. Vous pouvez aussi ajouter la description, par exemple. Ensuite, les trois autres grandes étapes. Eh bien, on a tout d'abord le répertoire des points de codes, les variantes et les règles contextuelles. Donc, on va passer en revue chacune de ces étapes. Mais avant cela, il faut aussi savoir qu'il y a certains principes de conception à prendre en considération au titre du RFC 6912, qui établissent les principes pour les points de code Unicode qui peuvent être inclus dans les étiquettes et utilisés dans le DNS.

Alors, il y a un bel éventail de principes. Par exemple, la longévité, c'est-à-dire lorsqu'on a de nouvelles lettres encodées dans l'Unicode. Eh bien, parfois, si cette lettre vient d'être encodée dans les versions plus récentes de l'Unicode, le logiciel ou l'appareil peut-être ne vont pas accepter cette nouvelle forme. Donc, même si la lettre a été encodée, peut-être qu'elle ne va pas être acceptée par les logiciels. Donc, ça veut dire que les points de code utilisés dans le système des noms de domaine doivent l'avoir été pendant déjà un petit temps pour nous assurer que c'est stable.

Il y a également le principe du moindre étonnement. Par exemple, si on utilise une langue ou un script bien particulier et qu'il y a un caractère qu'on ne voit jamais -- qu'on n'utilise jamais dans sa vie

au quotidien alors qu'elle apparait dans l'Unicode, eh bien, cela peut créer la confusion et peut-être que la communauté de ce script en question ne le comprendrait pas.

Il y a également le principe de sécurité contextuelle, c'est-à-dire que, parfois, il faut un contexte pour savoir comment utiliser le caractère, et, si c'est le cas, il faut que le contexte soit bien clair.

Il y a également un principe de conservatisme. Ce qui veut dire que tout le monde peut utiliser les ressources de ce système de nom de domaine à travers le monde, mais parfois, on ne sait pas si un point de code devrait être utilisé ou non. Aussi, si on n'est pas certain, il vaut mieux adopter une approche plus sure, donc être un peu plus conservateur dans son utilisation.

Ensuite, le principe de l'inclusion concernant tel ou tel code point qui devrait être utilisé. Donc, ça ne veut pas dire que -- en fait, au point de départ, on peut inclure absolument tous les points de code. Il faut aller petit à petit, procéder par étape, inclure d'abord ce qu'on peut et puis continuer à en inclure au fur et à mesure.

Et maintenant, je vais passer au dernier principe la stabilité. Si on autorise à ce qu'un point de code soit utilisé dans le nom de domaine, il faut s'assurer qu'il ne soit pas retiré par la suite.

Et donc voilà, tous ces différents principes doivent être pris en compte.

Alors, par rapport au répertoire des points de code, alors, quels sont les points de code qu'on peut utiliser pour un script en

particulier ? C'est la question. Quel est l'objectif derrière tout ça ? Eh bien, pour vous rappeler, l'idée, c'est de créer un système mnémotechnique pour qu'on puisse l'utiliser dans le système des noms de domaine. C'est ça, en fait, l'essence même du nom de domaine. L'idée, c'est de pouvoir se rappeler plus facilement des adresses IP. L'objectif, bien sûr aussi, est que ce système soit sûr et stable, qu'il ne doive pas prêter à confusion.

Et il ne faut pas non plus qu'un nom de domaine couvre forcément des exigences linguistiques ou la syntaxe grammaticale d'une langue. On peut utiliser, par exemple, des abréviations et l'orthographe ne doit pas forcément être correcte dans la langue ou le script utilisé. Par contre, ce nom de domaine doit être stable et prévisible.

Alors maintenant, revenons un petit peu sur les noms de domaine ASCII, donc du code standard américain pour l'échange d'informations. Alors, il y a peut-être des règles que vous ne connaissez pas.

Au premier niveau, avec les lettres ou le code ASCII, on ne peut utiliser que les lettres de A à Z. Pourquoi ? Eh bien, c'est parce que si on utilise des nombres aussi au premier niveau, cela pourrait prêter à confusion. Donc, au premier niveau, avec le code ASCII, on ne peut utiliser que des lettres. Mais au second ou troisième niveau, bien sûr, vous pouvez utiliser des lettres, des chiffres ou des tirets.

Il y a déjà donc des règles qui sont établies ici. Vous avez donc le tableau des points de codes pour le code ASCII, donc les lettres que

vous pouvez retrouver sur votre clavier d'ordinateur. Mais, outre ces lettres, il y a également des signes, des symboles mathématiques et d'autres commandes. Par exemple, la barre d'espace. Donc, ici, vous voyez souligné -- entouré en vert, ce sont les lettres que vous pouvez utiliser et seulement au premier domaine, au premier niveau. Mais, par contre, au deuxième niveau, vous pouvez utiliser les lettres, les tirets ou les chiffres.

Ensuite, ici, vous avez le tableau pour l'Unicode, le code Unicode, qui permet donc d'utiliser différentes lettres et plus de chiffres qu'avec ASCII. Donc ça, c'est-ce qu'on appelle le code Unicode.

Ici, vous avez à gauche un exemple pour le tableau pour l'arabe et à droite pour l'éthiopien.

Alors, nous en sommes maintenant à la version 17 de l'Unicode, qui couvre déjà 172 scripts. Donc, à gauche ici, vous avez ce que représente l'ASCII. En gros, on ne peut utiliser qu'un seul script et seuls 63 scripts sont couverts. Il y a déjà, par exemple, des confusions entre le zéro et le O ou d'autres lettres, alors que si on passe à droite sur l'Unicode, on couvre près de 300 000 caractères, ce qui peut prêter encore plus à confusion. Donc, il faut travailler avec la communauté pour identifier combien de points de code sont nécessaires, a minima pour couvrir les noms de domaine dans les différents scripts couverts par Unicode.

Alors nous avons commencé par passer -- donc comment est-ce qu'on fait ? Eh bien, on commence par passer en revue le tableau de code Unicode pour ce script, et on se demande si c'est valide.

Ensuite, on se demande si tel ou tel caractère est nécessaire, utilisé dans la langue ou dans le script. Est-ce qu'il est convenable ? Est-ce qu'il peut être utilisé pour les identifiants ? Par exemple, le fait de répéter un signe n'est pas forcément nécessaire dans un nom de domaine. Est-ce que le code est déjà stable ? Est-il accessible ? Est-ce que ces lettres sont, par exemple, représentées sur le clavier des utilisateurs ? Et puis on en revient au RFC 6912 quant aux principes régissant l'Unicode ; est-ce que ce code est en contradiction avec l'un des principes ? C'est les questions à se poser.

Passons maintenant à la section suivante, donc, concernant les variantes. Un autre point important des IDN, eh bien c'est qu'il y a des points de codes ou des ensembles de points de codes pour lesquels la communauté d'un script peut penser qu'ils sont les mêmes.

Par exemple -- et il y en a deux grands types. Tout d'abord, il y a des points de codes qui visuellement sont similaires. Et il faut bien les définir pour des raisons de sécurité. Ici, vous voyez, par exemple, le texte en bleu .EPIC en haut et à droite. Si vous regardez les points de codes qui se cachent derrière, qui traduisent ces noms de domaine, eh bien, le premier commence par 0065, le deuxième par 0435. Donc, il s'agit là en fait de deux scripts différents et de deux étiquettes différentes. Mais par contre, l'œil humain ne peut pas faire la différence entre ces deux noms de domaine. Et donc c'est pour l'alphabet latin et le cyrillique. Et donc, pour ces deux types de noms de domaine, il faut les définir comme étant des variantes. Et la communauté du script latin et la communauté du script

cyrillique ont travaillé ensemble justement là-dessus et ont défini ce 0065 et ce 0435.

Il y a également d'autres types de variantes, qui sont les variantes de d'usabilité, qui s'appliquent en majorité d'utilisations. Et cela s'applique, par exemple, aux scripts chinois. Vous voyez ici. Le deuxième caractère est la version simplifiée du chinois par rapport au deuxième caractère, qui est la version du chinois traditionnel. Ces deux caractères peuvent être utilisés dans des régions différentes du monde. Donc, si vous êtes un titulaire de nom de domaine ou de marque, que vous êtes présent en Chine, mais que vous voulez être présent dans d'autres parties du monde, vous voudrez peut-être avoir accès aux deux noms de domaine. Et dans ce cas-là, la communauté définit cela comme des variantes de ces derniers points de code. Et c'est la même chose pour le script arabe que vous voyez à l'écran.

Donc, la définition des variants est quelque chose qui doit être fait pour pouvoir créer des règles. L'objectif, c'est de garantir que ce système mnémotechnique minimise la confusion, donc de pouvoir nous assurer que les étiquettes fonctionnent de la même manière au sein d'un seul script, mais également entre les scripts. Ce conservatisme requiert donc des variants bloqués, qui sont donc garantis -- pour qu'ils garantissent la sécurité. Donc, si l'un des variants existe, cela veut dire que dans ce cas-là des variants bloqués, l'autre ne peut pas exister.

Dans un autre cas de figure, on veut pouvoir avoir ces variants actifs au même moment et, dans ce cas-là, ce sont des variants attribuables.

Alors, encore une fois, ce n'est pas une décision à prendre à la légère. Parfois, il faut discuter avec la communauté et également discuter avec d'autres communautés pour trouver ces conclusions.

Alors, pour définir des variants, il faut se poser les questions suivantes. Est-ce qu'il y a un type de points de code qui sont interchangeables pour les personnes qui sont natives du script qu'ils utilisent ? Est-ce qu'il y a une représentation déjà présente de ce variant ? Est-ce qu'il y a des variants de ce code qui sont pertinents au niveau du contexte ? Est-ce qu'il y a des variants qui rentrent en contradiction avec des principes ? Est-ce qu'il y a des variants interscript.

Alors, dans certains scripts, il peut y avoir des restrictions, par exemple sur l'IFC par rapport à la nature du script. Il faut s'assurer qu'il soit inscrit de manière sure et stable dans le DNS. Alors, encore une fois, il est important de se rappeler que nous ne suivons pas les règles grammaticales. Il faut s'assurer que le script n'est pas quelque chose auquel l'utilisateur ne pourrait pas s'attendre. Il ne faut pas également que cela crée des enjeux sécuritaires ou des contraintes d'utilisation. Et parfois, il y a certaines sauvegardes qui sont mises en place pour réduire les étiquettes attribuables et pour les rendre bloquées dans certains cas.

Alors, par exemple, nous avons des activités d'atténuation des risques. Par exemple, ces deux points de codes à droite et à gauche vous sembleront très similaires. On dirait des lignes verticales avec un petit point en bas. À gauche, c'est le point de code U plus 0E41. Mais à droite, ce sont en réalité deux points de codes similaires. Le point U plus 0E40. Dans ce cas, si vous voyez cela dans un contexte différent, donc dans une ligne de code, par exemple, ou sur un nom de domaine, vous pourriez confondre les deux. Donc, ce que nous avons fait, c'est de ne pas permettre que les deux points U et 040 puissent être utilisés l'un après l'autre.

Il y a également d'autres types d'atténuation des risques qui sont faits, par exemple sur les accents. Donc, la règle qui a été définie, c'est qu'une consonne ne peut avoir qu'un seul accent. Donc, si vous souhaitez écrire deux accents ensemble, certaines parties du système ne vont pas vous permettre de le faire, et cela dépend des systèmes. Par exemple, sur certains iPhones, cela peut être le cas et, sur certains Androids, ce n'est pas le cas. Donc, cette inconstance parfois peut créer la surprise chez les utilisateurs. Et c'est pour cela que, dans ce cas-là, par exemple, à l'écran, le script là-haut, il y a une règle très claire qui limite à un accent par consonne.

Donc, cette règle LGR, également, donc, des règles pour la génération d'étiquettes est utilisée dans les machines et est déjà présente avant la création des étiquettes pour pouvoir permettre la création de variantes de cette étiquette.

Voilà, j'en arrive à la fin de ma section. Je voulais vous présenter deux projets qui ont trait à la zone racine, aux règles pour la génération d'étiquettes de la zone racine.

Tout d'abord, la définition des communautés des scripts de différents panels, nous les avons intégrés dans la collection des LGR de la zone racine. Et désormais, nous avons 27 scripts qui représentent plus de 387 langues. Cela représente 11 ans de travail, 18 panels de générations, de personnes qui viennent de 45 pays. Et cela représente le travail de plus de 270 volontaires, ainsi que plus de 10 000 heures de travail.

Voilà donc les scripts qui sont disponibles grâce à cette zone racine, à ces règles de la zone racine.

Un autre projet sur lequel nous travaillons, ce sont les références LGR pour le deuxième niveau. Parce que si vous avez un gTLD ou n'importe quel type de TLD, vous voudrez peut-être développer le second niveau. Et ces règles doivent être révisées par l'ICANN. Donc, nous avons créé ces règles de référence pour pouvoir réviser ces scripts. Donc, nous avons désormais des règles pour les langues et des règles pour les scripts.

La collaboration pour développer ces références a été faite de manière intercommunautaire, encore une fois sur la base du volontariat. Et cette fois, nous avions besoin d'experts linguistiques pour développer le répertoire des variants et les

règles. Nous avons également demandé à la communauté de nous donner des listes pour pouvoir faire une analyse plus poussée.

Du point de vue de l'ICANN, nous avons obtenu du soutien technique sur les règles liées aux étiquettes et enfin une révision finale. Une fois que cela sera finalisé, l'ICANN org lancera une période de commentaires publics pour que cela soit validé par la communauté. Donc, si vous êtes intéressés par l'ajout de votre langue ou de votre script, vous pouvez rentrer en contact avec nous et nous pourrons lancer cela.

Donc, je vais donner la parole à Akshat Joshi, qui travaillait avec nous sur ce projet de LGR, sur le script du nouveau Brahmi, et comment nous pouvons inscrire cela dans l'aspect technologique.

AKSHAT JOSHI Merci beaucoup, Pitinan. Nous avons donc vu toutes les langues présentes dans ces règles pour la génération d'étiquettes et quel est l'objectif de cette LGR, de ces règles de génération d'étiquette. Alors je voulais vous présenter à quoi ressemblent ces règles. Comment -- si vous essayez d'utiliser ces règles, comment vous pouvez le faire pour essayer de voir quelles langues vous pouvez ajouter ?

Il y a quatre sections comme Pitinan l'a déjà expliqué. Il y a donc les métadonnées. Je récapitule parce que, pour la prochaine diapositive, vous allez voir que c'est assez important de se rappeler de cela. Nous avons donc la section métadonnées, la section

répertoire des points de code, la section des variants et de leurs informations, et enfin nous avons les règles qui sont des règles de contexte, des règles d'évaluation et des règles d'action.

Ces règles sont présentes dans deux formats différents. Donc, nous avons le format XML et le format HTML. Et vous pouvez utiliser des outils LGR pour générer l'équivalent XML ou HTML.

Alors, cet outil LGR est en réalité fourni par l'ICANN, qui permet de développer et de gérer les règles. Donc, si vous voulez faire valider votre script par ces règles, vous pouvez utiliser cet outil. Et si vous souhaitez vérifier avec les tableaux IDN qui existent, c'est également possible grâce à cet outil.

Alors, comment vous pouvez accéder à cet outil ? Eh bien, c'est simple. Vous pouvez taper lgrtool.icann.org, et, grâce à cela, vous pouvez déjà utiliser cet outil.

Il y a également des guides d'utilisateur qui sont disponibles sur la page Web qui est liée à cette diapositive. Cet outil, c'est un outil à code source et libre qui peut vous intéresser.

Alors, pour créer ces règles, nous allons inventer une langue. La langue A à G. Alors, quelles sont les caractéristiques de cette langue ? Donc, nous l'inventons, c'est une langue fictive qui serait une langue de script latin avec seulement 11 lettres et un signe. Donc, voici les lettres qui sont utilisées par les personnes qui parlent cette langue.

Les prérequis. Les exigences de cette langue, c'est qu'il y a deux types de caractère F ; donc nous avons un F normal et le F hameçon. Et la communauté considère les deux comme interchangeables. Donc, si on veut écrire « fin », on peut utiliser le F normal ou le F hameçon. Donc, il faut pouvoir utiliser les deux versions, les deux variantes.

Donc, si on regarde le tableau d'Unicode, nous voyons qu'il y a un autre code, le clic alvéolaire, qui est utilisé dans le latin basique. Et donc nous avons trois caractères latins. Donc, si on regarde les répertoires de langue, on a 11 caractères latins basiques, 3 suppléments, et nous avons également 2 caractères du latin étendu B. Donc, ce petit F hameçon dont je vous parlais est ce clic alvéolaire.

Vous voyez qu'à l'écran il y a un carré rouge. Et je vais vous expliquer pourquoi. Alors, ce signe du degré ne peut pas être permis. Alors, vous vous rappelez de cette exigence avec le caractère F. La communauté les considère comme interchangeables. Donc, il faut pouvoir créer des variants avec ces deux caractères. La disposition doit être de type bloqué. Donc, si la communauté utilise ce type de F, le deuxième F ne peut pas être présent dans le système.

Il y a également une autre règle -- une autre exigence, c'est que le clic alvéolaire ne peut pas être au début d'une étiquette. Donc, nous allons voir comment cela se présente dans le format LGR. Vous voyez donc sur cette diapositive deux versions des règles.

Vous avez à gauche la version XML. Et à droite, nous avons la version HTML. Donc, la version XML, c'est une version utilisée par la machine, et, HTML, elle est plus lisible pour l'œil humain. Donc, si on regarde la version HTML, vous voyez ce dont nous avons parlé, toutes les informations qui sont présentes. Il y a également un élément de données d'informations. Et vous avez les variantes. Donc, vous avez le variant le 0066 et le 0192. Donc, vous voyez, il y a un principe de symétrie ici que, si un caractère est présent, le deuxième doit être le variant de l'autre.

Et pour la dernière section, qui est donc le dernier élément, c'est la section sur toute l'étiquette. Donc vous avez le nom au début du mot et ça s'applique à ce qui est alvéolaire.

Diapo suivante, s'il vous plaît.

Alors, maintenant que nous avons créé cette LGR, eh bien, une des fonctions de cet outil LGR dont on a parlé, c'était de voir si un caractère est valable ou pas. Donc, ici, l'idée, c'est d'approuver ou non des étiquettes qui pourraient être utilisées par la communauté.

Donc, tout d'abord, on a l'étiquette ABÉ avec un accent aigu, et cela, c'est tout à fait valide, car tous les points de code proviennent du répertoire et il n'y a pas de règles particulières à appliquer. Donc ça, c'est une étiquette totalement valable.

Ensuite, deuxième étiquette, donc ÄBÉ avec l'accent aigu. Par contre, il y a un A avec un tréma. Et ce caractère-là ne fait pas partie

du répertoire de cette langue AG. Donc, cette étiquette n'est pas valide pour cette langue.

Ensuite, on en arrive à la question des variants. Rappelez-vous donc on a le F normal et le F hameçon. Alors, première étiquette F à accent aigu. Cette étiquette est valable, mais il y a une information supplémentaire ; c'est-à-dire que cette étiquette a une variante qui est bloquée, c'est-à-dire tous les caractères sont les mêmes, sauf le F, qui est un F hameçon. Et donc nous avons cartographié cette variante comme étant bloquée. Voilà pourquoi le statut ici est bloqué.

Ensuite, la deuxième étiquette. Et donc le mot fi avec un F hameçon qui est une étiquette valable. Et on constate encore une fois que l'étiquette variante fi avec un f normal est bloquée. Donc ça veut dire que si un mot est écrit avec un des deux F de ce de cette langue, la variante avec l'autre F sera bloquée.

Ensuite, on passe à d'autres étiquettes. Rappelez-vous donc la règle concernant le clic alvéolaire qui ne peut pas se retrouver au début de l'étiquette. Ici, le CA clic, eh bien, c'est une étiquette tout à fait valide. Par contre, la deuxième étiquette commence avec ce clic alvéolaire, raison pour laquelle cette étiquette n'est pas valide. Et alors là, vous avez une description expliquant pourquoi cette étiquette n'est pas valide. Et donc vous voyez la règle LGR.

Diapo suivante. Alors maintenant, prenons un exemple de caractère similaire pour une étiquette *devanagari*. Ici, une erreur a été faite délibérément. Le signe *matra* que vous voyez ici à l'écran

est repris deux fois, alors que, en fait, il ne devrait être repris qu'une seule fois, raison pour laquelle cette étiquette est invalide. Et encore une fois, on vous explique quelle est la raison pour laquelle ce n'est pas valide.

Ensuite, ici, vous avez donc un exemple intéressant de variantes, donc entre deux scripts différents. Donc, à gauche, vous avez une étiquette du script *devanagari*. Et à droite, vous avez donc l'exemple de l'étiquette *gurmukhi* -- du script *gurmukhi*. Alors, les utilisateurs du *devanagari* pourraient dire que ces deux étiquettes sont similaires, qu'elles se ressemblent, raison pour laquelle on les a encodées comme étant des variantes.

Voilà, c'en est tout pour moi.

PITINAN KOOARMORNPATANA Merci beaucoup à tous. Maintenant, je voudrais donner la parole à Louis afin qu'il nous fasse part des informations concernant les LGR pour les langues des Premières Nations.

LOUIS HOULE Merci beaucoup. Alors si vous ne me connaissez pas, je m'appelle Louis Houle. Je suis coprésident du groupe de travail UCAS, donc du syllabaire autochtone canadien unifié. Aujourd'hui, je serai très bref.

Tout d'abord, je voudrais tout d'abord vous demander de vous pencher sur cette carte que vous voyez à l'écran. En bleu, vous voyez les dialectes du peuple inuit, qui sont utilisés au Canada --

dans le nord-est du Canada. En orange, vous voyez donc les scripts des *Cree*. Vous avez donc à droite, en orange, les dialectes des *Cree* de l'est et de l'Ouest, et en rose les scripts d'un autre peuple. Donc, il y a 21 langues différentes de script syllabique, tandis qu'il y a 80 langues de script latin. Ce qui veut dire qu'il y a donc 101 langues parlées au Canada. Donc, c'est un nombre assez impressionnant de langues différentes.

Diapo suivante. Alors, quel est le champ d'application de notre travail ? Dans le contexte de l'ASCII, donc du code standard américain pour l'échange d'informations et des noms de domaine de second niveau français. Vous avez ici un exemple avec le français du Québec. Donc, en fait, les Premières Nations utilisent de manière assez égale l'anglais ou le français pour disposer de noms de domaine, mais, par contre, ils n'ont pas de nom de domaine dans leur propre langue. Par exemple, il n'y a pas le naskapi.ca ou .QUEBEC. On n'a pas le *cree* de l'Est ou .CA ou peu importe le code inuktitut ou Inuit. Voilà.

Donc, il y a des groupes de travail sur les scripts syllabiques et latins. L'idée est donc de documenter le répertoire des langues locales et de les établir avec des points de code Unicode, et de valider aussi les caractéristiques des LGR proposées pour ces scripts. Pour le moment, bon, le système dont on dispose n'est pas très bon. Il est un peu obsolète.

Ensuite, il faut aussi évaluer l'utilisation à l'heure actuelle de certaines langues ou de certains caractères, de certains mots. Il y a

parfois des nouveaux mots qui apparaissent au sein d'une langue et qui n'ont pas été encodés par le passé. Donc, nous définissons aussi les variantes au sein d'un même script ou alors entre différents scripts.

Nous définissons des règles. Nous sommes en train de finaliser notre proposition. Et Madame la Présidente de cette séance nous a beaucoup aidés pour tout ce qui était format XML ou HTML. Nous sommes prêts maintenant à passer à la période du commentaire public et l'intégration sera l'étape suivante.

Diapo suivante, s'il vous plaît. Alors, en ce qui concerne les langues de script syllabiques, eh bien, nous avons une LGR pour la langue *inuktitut* qui couvre cette sous-langue. Ce document a été publié en 2025 et cette LGR donc maintenant a été finalisée. Elle comprend *l'inuktitut*, le *cree* et *l'ojibwé*. Cela représente donc 18 sous-langues ou dialectes. Et le commentaire public sera disponible en mars-avril 2026 pour la LGR, donc, pour les scripts *dene* et *dakelh*. Là, la période de commentaires publics sera prévue pour juillet-septembre 2026.

Ensuite, pour les langues de script latin, nous avons finalisé donc une LGR qui couvre les langues *nunavut* et *l'inuktitut* de la région du Labrador, de Québec et des provinces maritimes Et nous avons également complété, figurez-vous, hier, la LGR pour les langues en Ontario. Il reste encore des LGR à rédiger pour une langue de script syllabique et pour 48 langues de script latin.

Diapo suivante. Je voudrais rebondir sur ce que vous avez dit et vous donner un exemple de ce sur quoi nous travaillons.

Nous visons aussi, bien sûr, à réduire tout risque potentiel concernant les homoglyphes, les tentatives d'hameçonnage, de détournement ou de violation du droit des marques. Et ici, nous nous concentrons sur des mots de second niveau et non pas d'extensions de premier niveau.

Exemple de langage de script latin. Nous avons le .QUEBEC. Et en français, il y a tout un éventail de possibilités pour cet E avec différents accents. Et pour tout nom que vous allez choisir, vous avez deux possibilités qui s'offrent à vous : un non sous format ASCII et un mot sous format français avec, vous voyez donc, les différents accents que vous pouvez utiliser. Et, en fait, on peut combiner les deux. Vous avez donc besoin d'un outil qui vous permet de le faire, parce qu'il y a aussi tous les signes diacritiques que vous pouvez voir à droite en vert, que vous pouvez aussi utiliser.

Alors, il faut remarquer que certains des mots qui existent disposent peut-être donc d'un diacritique qu'on pourrait autoriser, mais qu'on ne veut pas autoriser, parce que cela va juste prêter à confusion parmi les utilisateurs. Donc, bien évidemment, nous nous penchons aussi sur la question de l'intérêt public et de la sécurité du DNS.

Diapo suivante. À l'heure actuelle, nous en arrivons à la fin de nos travaux et nous espérons que cela aura pour conséquence que

l'Internet sera encore plus ouvert. Nous aurons des réunions lors des semaines à venir, et nous espérons vraiment que les résultats de notre travail permettront que l'Internet soit accessible à plus de personnes, qu'il soit plus accessible aux Inuits et aux Premières Nations au Canada. Et, bon, comme vous le savez, l'Internet, c'est un bien public pour tout le monde, comme on dit « Un monde, un Internet ». Merci.

PITINAN KOOARMORNPATANA Merci beaucoup, Louis, pour ces informations concernant le Canada. Maintenant, je voudrais donner la parole à Gheafany d'Indonésie, qui s'exprime à distance et qui va nous parler des LGR de référence pour les scripts indonésiens. Gheafany, vous avez la parole.

GHEAFANY WIDYATIKA PUTRI Merci beaucoup, Pitinan. Bonjour à toutes et à tous. Je suis Gheafany de .ID et je vais vous donner des informations sur les LGR pour notre région, notamment pour les scripts balinais, javanais et pégon.

Alors, un nom de domaine, c'est notre adresse numérique. Et si cette adresse IP ne peut être écrite que dans un script, alors l'Internet ne parle pas votre langue. En indonésien, on utilise des scripts locaux dans différents domaines. Aussi, les IDN permettent aux gens d'utiliser leur propre script en ligne.

[Alors malheureusement, le son de Gheafanyva et vient. L'interprète fait de son mieux.]

Notre mission est donc de permettre une plus grande inclusion, tout en protégeant les utilisateurs qui sont exposés à des risques. C'est notamment l'essence de notre travail.

Diapo suivante, s'il vous plaît. À un haut niveau, les LGR nous permettent de faire beaucoup de choses, comme Pitinan l'a déjà expliqué. Tout d'abord, l'idée, c'est de pouvoir utiliser différents caractères, de se pencher sur ce qui fonctionne en pratique, et d'assurer aussi la sécurité. Alors, bien sûr, tout ce travail n'est pas fait par une seule communauté, mais par toute une -- par un seul groupe de personnes, mais par toute une communauté plutôt composée d'experts, de différentes parties prenantes, de chercheurs académiques, etc., etc. Nous travaillons sur la question de la prévention et nous travaillons aussi à ce qui est acceptable ou pas. Et nous avons des chercheurs académiques ou des praticiens qui travaillent à ces questions. Vous avez ces personnes à l'écran et [Pandy] se fait l'agent de liaison avec l'ICANN. Et donc, tout ceci est une pratique communautaire. Et en réalité, sans ces points focaux, sans ces gardiens, je pense que nous ne pourrions jamais atteindre le consensus.

Prochaine diapositive, s'il vous plaît. Alors je voulais maintenant vous parler des progrès que nous avons menés, tout d'abord, sur le script balinais. Le balinais est un jalon pour l'Indonésie. C'était le premier script à atteindre la référence LGR de second niveau à

l'ICANN en novembre 2024. Cela a été lancé avec le gouverneur de Bali en décembre 2025.

Le deuxième script est le javanais. Pour cela, l'ICANN a visité Yogyakarta pour évaluer l'utilisation du javanais directement avec les parties prenantes. Nous sommes désormais en phase de finalisation avec l'ICANN et nous sommes bientôt à la phase de publication à la suite de la phase de commentaires publics.

Et encore une fois, je vous rappelle que nous ne créons pas de règles, mais nous regardons l'utilisation réelle des scripts. Le dernier script, c'est le script pégon. La référence LGR a été soumise et nous attendons le retour de l'ICANN. La prochaine étape sera d'incorporer ces retours pour nous assurer que ce script soit utilisable de manière sure.

Donc, voici un cas pratique pour vous montrer l'équilibre entre l'utilisation et la sécurité. Les noms de domaine sont toujours inscrits à l'écran et parfois en fonction de l'écriture, il peut y avoir des confusions. Par exemple, à l'écran, vous avez le 1 javanais et la lettre go javanaise. Donc, le un est très similaire à cette lettre, donc les utilisateurs peuvent -- cela peut prêter à confusion.

Alors, ceci est très important à prendre en compte lors de l'élaboration des politiques également. Donc, si vous voulez créer un domaine comme ICANN1, vous pouvez écrire ICANNone en anglais ou 1 pour un. Nous mettons donc en œuvre des pratiques

réelles d'écriture de noms de domaine pour pouvoir garantir l'utilisation par les utilisateurs finaux.

[L'interprète s'excuse. Le son de Gheafany est très instable.]

Alors, pour ce qui nous attend pour la suite, nous allons compléter la finalisation de la référence LGR pour le javanais. Nous allons passer à la phase de commentaires publics. Et pour le pégon, nous allons évaluer les commentaires de l'ICANN. Et nous avons hâte de faire cette avancée pour d'autres scripts indonésiens, comme le batak, le buguinais ou le soudanais en fonction de la préparation de la communauté et de la demande.

PITINAN KOOARMORNPATANA Merci beaucoup Ghea. Voilà donc la fin de notre présentation. Est-ce que vous avez des questions ou des commentaires ? Et avant de prendre la parole, n'oubliez pas d'annoncer votre nom.

DIPAK PARMAR Merci beaucoup ! J'avais une question assez basique à vous poser. Comment est-ce que l'utilisateur final tape le nom de domaine ? Parce qu'il y a une vraie différence au niveau de la contribution. Mon anglais est -- mon clavier est différent en fonction du pays d'où je viens. Est-ce que cela prend en compte [dans] les contributions de la communauté ? Parce qu'il faut prendre toutes ces considérations en compte.

LOUIS HOULE Je peux parler de notre expérience avec les Premières Nations. Nous avons des personnes, au sein du groupe de travail, qui sont des spécialistes des claviers. Et ils nous ont aidés à effacer tous les problèmes qui existent toujours encore aujourd'hui, avec certains claviers qui sont utilisés, qui ont des caractères qui sont obsolètes ou qui ne sont plus utilisés. Et je ne parle pas que d'une personne ici, je parle de l'autorité des claviers dans le monde. Donc, les claviers UCAS n'existent pas encore.

DIPAK PARMAR Alors, vos experts ne sont peut-être pas des utilisateurs finaux ou des experts.

PITINAN KOOARMORNPATANA Alors, le projet que nous menons est un projet de nom de domaine. Et la manière de développer ces noms de domaine, c'est d'abord une manière visuelle. Donc, d'avoir une contribution liée au clavier est toujours important. Mais nous restons sur une communauté assez technique. La règle actuelle est toujours une règle visuelle et les communautés techniques ont des règles liées aux variants qui ne s'étendent pas encore au niveau des accents et des claviers, mais merci pour votre contribution. Il y a toujours une question sémantique, mais merci pour votre contribution.

MOHAMMAD ZULQAR NAYEN Zulqar au micro. Merci. J'ai une question rapide à poser à Louis.

Le travail que vous faites avec les Premières Nations au Canada et leur syllabaire autochtone, est-ce que les communautés autochtones sont les personnes qui ont fait la demande de ces noms de domaine ou est-ce qu'ils ont rejoint le travail ?

LOUIS HOULE

Eh bien, les deux. En réalité, cela dépend de la langue. Cela dépend de la région. Ce que nous avions voulu faire au départ, c'était assez simple, mais c'était peut-être limité au début. Ce que nous voulions offrir avec le .QUEBEC -- l'opérateur de registre Québec, c'était d'avoir différentes langues. On a commencé avec *l'inuit*. Et en réalité, sur les 27 langues, il n'y en avait que deux qui étaient parlées au Québec.

Si je prends toutes ces langues, encore une fois, on a le *cree*. On a *l'ojibwé*. Le problème, c'est que, au Canada, comme vous le savez, les Premières Nations, et vous l'avez vu sur ma première diapositive, ce sont -- j'ai mis des cercles. On a le Québec, qui est tout petit à l'est, mais on a des Premières nations qui sont présentes sur des territoires bien plus vastes que le Québec. Donc, le concept lui-même de Québec, d'Ontario, de territoire maritime de Colombie-Britannique, ce n'est pas quelque chose que l'on peut utiliser avec les Premières Nations parce que ce ne sont pas leurs territoires. Cela ne s'applique pas. Donc, on vient seulement de découvrir quelque chose qui était évident pour la plupart de ces utilisateurs, et nous avons été peut-être trop naïfs. C'est qu'une fois qu'on a commencé, il est très difficile de s'arrêter.

LANCE HINDS

Merci, madame la Présidente. Je suis le président de LACRALO et j'avais un groupe de travail qui portait sur l'utilisation de la technologie pour préserver les langues dans les Caraïbes.

En ce qui concerne les Premières Nations et les consultations qui ont eu lieu, je suis très curieux de savoir ce qui s'est passé sur le terrain parce que nous n'avons pas ces compétences. Donc, nous avons un vrai manque de documents et parfois nous avons dû évaluer les langues pour comprendre quels sont les diacritiques utilisés. Et parfois, il y a des conflits. On a l'impression qu'il y a donc un guillemet opposé ou parfois une apostrophe. Cela peut être confondu. Et je suis assez curieux d'entendre votre expérience là-dessus.

Donc, la règle, c'est de se référer à la communauté parce que c'est elle qui va définir quel symbole diacritique est le plus utilisé. Et le problème des bonnes pratiques, c'est qu'elles ne sont pas universelles et on ne peut pas les utiliser seules. Donc, que se passe-t-il lorsqu'on a ce type de symboles qui peuvent être confondus -- ce type de conflit ? Et comment est-ce que vous gérez cela ?

PITINAN KOOARMORNPATANA

Merci beaucoup pour la question, Je crois que je vais commencer par y répondre et Louis aura un bon exemple.

Alors, quand on revient à ces étiquettes, il faut s'assurer qu'elles soient conformes à l'IDN, donc aux noms de domaine internationalisés dans les applications. Donc, si on revient sur ces guillemets simples ou le son, l'interruption [inaudible] de nasales dans certaines langues, cela fait partie du protocole de faire cette différence.

Et voilà le statut actuel. Certains de ces symboles ne sont pas encore présents dans le PVALID. Je pense que c'est quelque chose que Louis peut partager également sur son expérience.

LOUIS HOULE

Oui, alors on utilise PVALID et on a essayé de voir ce qui était possible tellement de fois sur ces questions. Alors encore une fois, je ne suis pas spécialiste de toutes les langues des Premières Nations. Moi, je suis spécialiste du français.

Et pour vous dire, dans notre équipe, nous avons des personnes qui peuvent répondre à ces questions. Nous avons des personnes qui ont publié le dictionnaire *naskapi* en 2025, et donc nous pensons que c'est un expert qui s'y connait dans cette langue.

Et nous sommes toujours en lien avec la communauté. Dès que nous avons une question, un problème, que ce soit valide ou non valide, la question se pose toujours de l'utilisation. Est-ce que ce caractère est appris à l'école ? Est-ce qu'il est utilisé au quotidien ? Et nous ne sommes pas les décideurs ultimes. C'est la communauté qui décide. C'est l'utilisation qui prime.

PITINAN KOOARMORNPATANA Merci beaucoup. Nous avons une main levée. Bonjour. Vous avez la parole, Bill.

BILL JOURIS L'un des problèmes que nous voyons vient en réalité du fait que dans les projets IDN, et même encore plus, les membres de l'ETF — du groupe de génie Internet — qui ont écrit les normes pour ces IDN, eh bien, aucune de ces personnes n'était natif de ces clics, de ce type d'arrêt glottal. Et ce sont peut-être des personnes qui ignoraient même cette possibilité d'utiliser des points de ponctuation pour les représenter.

Donc, en réalité, dans ce cas-là, les personnes qui ont écrit les règles n'ont même pas pensé que c'était important. Et c'est pour cela que l'on utilise l'apostrophe comme coup de glotte. Tout simplement, parce que les personnes qui ont écrit les règles n'en avaient cure. Et je pense que c'est quelque chose qu'il faut changer. Il faut le prendre en compte. Merci.

PITINAN KOOARMORNPATANA Merci Bill. Merci pour le commentaire. Alors, pour rebondir là-dessus, il y a également plusieurs points à prendre en compte. Par exemple, dans certains scripts, il y a des lettres mêlées qui peuvent être utilisées tous les jours, mais qui ne sont pas forcément visibles dans les noms de domaine de premier niveau, parce que ces deux lettres peuvent également être utilisées séparées. Donc, cela fait

partie des règles qui sont créées par l'IETF. Et évidemment, la demande prime. Donc, s'il y a une demande d'utiliser ce type de caractères, c'est à ce moment-là que nous chercherons des solutions ensemble. Mais merci, Bill, pour votre contribution.

CHAMPIKA WIJAYATUNGA Nous arrivons à la fin de cette séance. Il n'y a pas d'autres questions dans la boite de questions-réponses. La séance est donc maintenant levée. Merci à tous les participants et nous pouvons arrêter l'enregistrement.

[FIN DE LA TRANSCRIPTION]