

ICANN72 | Semaine de préparation – Point sur les règles de génération d'étiquettes pour la zone racine
Jeudi 14 octobre 2021 – 11h00 à 12h00 PDT

PITINAN KOOARMORNPATANA: Bonjour à toutes et à tous. Nous allons maintenant lancer l'enregistrement. L'enregistrement est lancé. Merci.

Bonjour ou bonsoir à toutes et à tous et bienvenue à cette séance de mise à jour sur les règles de génération d'étiquette de la zone racine. Nous avons l'ordre du jour. Je vais passer rapidement pour vous donner quelques informations sur les règles de génération d'étiquette. Nous aurons nos intervenants, Mats Dufberg qui nous parlera de l'écriture latine, Hiro Hotta du Japon qui nous parlera des scripts en japonais et enfin Michel Suignard nous parlera de la prochaine version des règles de génération d'étiquette.

Nous allons passer à la première partie : historique rapide de ce projet. J'espère que vous pouvez bien le voir.

Environ en 2010, la communauté de l'ICANN a identifié un besoin pour avoir des domaines de premier niveau et des variantes sur ces domaines de premier niveau. Nous avons eu le rapport intégré qui a identifié ce besoin de définir des TLD variantes. Nous avons exploré si nous allions pouvoir mettre sur la zone racine ces variantes, donc c'est ce qui a été étudié. On n'a pas trouvé de solution à l'époque.

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.

Dans ce rapport intégré, il y avait six groupes, le cyrillique, le grec, le chinois... Nous avons eu plusieurs types d'écritures et de scripts qui ont été analysés. Le rapport a été développé et publié en 2012. Et on a identifié des mécanismes pour ces variantes dans ces différents types d'écriture. Et nous devrions avoir une seule source pour les TLD pour la zone racine.

Ensuite, on a identifié des procédures de règles de génération d'étiquette LGR qui ont été approuvées par le Conseil d'Administration de l'ICANN pour leur mise en œuvre avec les gTLD. Il y avait donc des panels qui ont été créés pour analyser les différents types d'écriture. ICANN Org a également effectué des recommandations pour la gestion des différentes variantes des TLD pour les IDN, les noms de domaine internationalisés, et cela a été approuvé en 2019. Il a été résolu que la GNSO et la ccNSO prennent en compte les recommandations pour la gestion de ces TLD et cela, pour l'intégration de l'utilisation des LGR dans la zone racine.

En 2020, le Conseil d'Administration a également décidé que la GNSO et la ccNSO devraient prendre en compte les recommandations pour l'utilisation technique de la génération d'étiquette et de ses règles pour la zone racine. En 2021, la GNSO a publié son rapport sur les SubPro, les procédures ultérieures, qui incorporait l'utilisation de ces règles de génération d'étiquette pour la prochaine série de nouveaux gTLD.

Donc nous avons comme une formule, si vous le voulez. Et lorsque nous mettons ces règles et que nous les intégrons dans la zone racine, cela sera lisible par les machines, par les outils informatiques. Et on

pourra le faire d'une manière mécanique. On pourra identifier différentes étiquettes et identifier des variantes également.

Le mécanisme et le processus de développement pour ces LGR dans la zone racine fonctionnent en deux étapes. Tout d'abord, le travail du panel sur ces étiquettes et sur ces règles par système d'écriture, par script, c'est donc trouver des solutions, de les proposer au panel d'intégration, qui va ensuite considérer les propositions du panel de génération. Et il peut y avoir plusieurs consultations entre le panel de génération et le panel d'intégration, processus d'acceptation pré-analyse, et des règles sont donc définies pour les différents panels, ce qui devient validé pour les différents TLD.

Voici un résumé du travail qui a été fait par le panel de génération. Pour 28 types d'écriture, nous avons un travail qui est effectué pour au moins 26. En 2014, nous avons commencé avec la langue arabe et vous avez vu qu'après de nombreux efforts, certains groupes ont finalisé leur travail. Nous avons encore l'écriture latine et le japonais sur lequel il y a un travail qui a été effectué, l'écriture du Myanmar également va bientôt être analysée.

Voici le calendrier de développement pour ces LGR. Demande de mise en œuvre de panel de génération en 2013. Nous avons eu des scripts finalisés et ils ont été mis en œuvre dans la zone racine. Nous avons eu 18 systèmes d'écriture et nous nous attendons en 2022 à avoir effectué un travail supplémentaire sur 8 systèmes d'écriture, ce qui nous fait un total de 26 scripts.

Aujourd'hui, nous allons voir plus particulièrement l'écriture latine et le système d'écriture japonaise. Ils travailleront également dans des commentaires publics. Donc nous allons donner la parole à Mats Dufberg, qui est membre du groupe de génération d'étiquette pour l'écriture latine.

MATS DUFBERG :

J'espère que vous m'entendez bien. Nous allons passer à la prochaine diapositive.

Voilà ce dont je vais traiter. Nous allons parler de la proposition que nous avons pour les commentaires publics. Ce sont les points principaux.

Nous avons une proposition pour l'écriture latine et les commentaires seront acceptés jusqu'au 23 novembre 2021. Vous avez un lien à l'écran et nous encourageons toutes les personnes à regarder de près notre proposition, à l'analyser et nous prendrons en compte tous les commentaires et toutes les suggestions. Que ce soit des commentaires mineurs ou des commentaires importants, tout est bienvenu de notre part.

Nous avançons dans la présentation, qui va vous montrer ce que nous essayons de faire. Et j'espère que cela abaissera le seuil qui vous permettra de commenter et qui vous permettra de bien assimiler le travail effectué. Nous allons nous concentrer sur le document principal et ses annexes. Mais le document normatif, c'est un fichier XML pour ces générations d'étiquette, mais je ne vous vais pas aborder cela.

Les chapitres qui seront débattus dans cette présentation seront des chapitres courts avec des informations générales. Nous allons passer au chapitre 2 – le chapitre 1, ce sont des informations générales. Le chapitre 2 définit les limites des écritures dans le cadre de cette proposition, qui premièrement ne peut pas inclure tout caractère n'était pas inclus dans le répertoire maximal initial, le MSR. Le MSR est un sous-groupe du point de code des protocoles pour les noms de domaine internationalisés dans les applications, qui est lui-même un sous-ensemble d'Unicode.

Le MRS a été défini par le panel d'intégration. Durant le processus, il y a eu quelques caractères que nous avons proposés pour rajouts au panel de notre intégration et ces caractères ont été ajoutés dans le répertoire maximal initial, le MRS, qui est un document vivant en Unicode. Grâce à Unicode, on peut augmenter sa taille. Il y a donc quelques caractères qui ont été proposés et acceptés.

Dans le chapitre 4, c'est la méthodologie et le processus de développement pour notre groupe. Nous devrions savoir quels caractères utiliser et nous avons décidé d'utiliser ce qui est expliqué, qui est un niveau EGIDS. C'est le niveau d'éducation et une échelle de disruption intergénérationnelle étendue. Nous avons le niveau 5 de développement par exemple. Ce sont les langues et c'est le niveau d'utilisation des langues.

Vous avez cette échelle EGIDS, cette échelle de disruption intergénérationnelle. C'est pour pratiquement toutes les langues que l'on trouve dans le monde. On leur donne des niveaux d'utilisation en ce qui concerne l'éducation ou le niveau national, de 0 à 5. Il y a donc

différents niveaux. Le niveau 5 par exemple, c'est une langue qui est beaucoup utilisée. Il y a beaucoup d'autres niveaux que nous avons exclus pour que ce soit plus clair. Il y a le niveau 6 et plus.

Nous avons une liste de toutes les langues sélectionnées dans l'annexe B. Je n'ai pas le chiffre exact avec moi, mais cela représente beaucoup de langues. Pour chaque langue incluse, nous avons... Bill Jouris nous indique qu'il s'agit de 212 langues qui sont incluses. Nous avons le nom de la langue dans cette liste. Nous avons le code linguistique en trois lettres de l'ISO, l'Organisation internationale de normalisation. Et nous avons le niveau EGIDS dont nous avons parlé tout à l'heure. Les caractères sont ceux qui sont utilisés par les langues sélectionnées. Nous les avons identifiés. Nous devons décider quels caractères utiliser. L'idée derrière cela, avec ces 212 langues, c'est que nous devrions être capables de créer des noms de domaine de premier niveau dans ces langues. L'ensemble des caractères de chaque langue n'est pas documenté dans le rapport, mais ils peuvent être trouvés en référence pour chaque langue qui se trouve dans le chapitre 9. Les candidats pour ces variantes ont également été identifiés. Et nous allons passer à la diapositive suivante.

Le répertoire que nous avons pour l'écriture latine dans la proposition est basé sur les points de code Unicode. C'est ce qui nous limite et c'est ce que nous utilisons. Un point de code, c'est simplement un caractère, comme la lettre A par exemple. Nous pouvons également avoir une modification avec un signe de ponctuation par exemple. Vous voyez des exemples à l'écran, le G avec le tilde, qui est une combinaison d'une lettre et d'un signe de ponctuation.

Dans de nombreux cas, l'Unicode a des points de code précomposés dans lesquels le caractère de base est combiné avec un accent, comme par exemple le A avec accent aigu. C'est seulement un point de code. Les principes pour inclure ou ne pas inclure un caractère identifié dans une langue sont indiqués dans l'introduction du chapitre 5.

La proposition dans le cas de la liste latine comprend 218 caractères dont 197 sont à un point de code, donc simples, et 21 sont formés par une séquence de deux ou plus de points de code, donc un caractère de base et une marque ou deux. Pour chacun des caractères, il y a le point de code dont consiste le caractère et c'est soit un seul point de code, soit une séquence. C'est le langage et les langages qu'utilise cette écriture de caractères, ce n'est pas nécessairement toutes les langues, mais suffisamment de langue pour vérifier qu'il s'agit d'un caractère qui est utilisé pour une langue déterminée. Les références à l'alphabet de ces langues apparaissent également sur la liste. Revenez en arrière, voilà.

Dans le cas des caractères A à Z, nous n'avons pas de langue qui doive en témoigner étant donné que ce sont des caractères qui sont compris par défaut. Diapositive suivante.

Le répertoire représente l'une des principales parties du fichier XML. Le répertoire qui apparaît dans le chapitre 5 est dans l'ordre numérique par code de point de code. Mais dans un même répertoire, les caractères sont regroupés en fonction de la forme de glyphe, donc les différents accents et les diacritiques. Tout cela apparaît à l'annexe

C. Ce sont les mêmes informations, mais autrement organisées. Diapositive suivante.

Dans la partie 4 du chapitre 5, nous avons énuméré les caractères exclus, à savoir les caractères qui sont utilisés dans au moins une langue sélectionnée, mais qui ne peuvent pas être inclus étant donné qu'ils n'appartiennent pas au MSR. Et par MSR, j'entends le processus préalable qui définit la limite supérieure de ce qui peut être inclus.

Le panel de génération latin ne peut pas inclure des caractères qui n'apparaissent pas sur ce processus MSR. Donc si vous trouvez une lettre qui vous semble devrait être incluse mais dont les contenus n'apparaissent pas, le panel de génération latin ne peut rien faire pour changer cela. D'abord, c'est le MSR qui doit l'inclure. Diapositive suivante.

J'ai déjà parlé du répertoire. Et la partie suivante de notre présentation portera sur les ensembles de variantes qui sont une partie importante de la proposition. Cela est expliqué dans le chapitre 6.

Alors, qu'est-ce qu'une variante ? Un ensemble de variantes consiste à deux caractères ou plus qui, d'une manière ou d'une autre, sont perçus comme étant le même. Ils sont traités comme s'ils étaient le même caractère dans un certain sens, donc ce peut être la même forme, soit parce qu'ils peuvent être utilisés de manière interchangeable pour une partie de la communauté du script ou pour l'ensemble de la communauté qui utilise ce script. Lorsque vous avez en ensemble de variantes, vous ne pourrez pas utiliser les deux

caractères de manière indépendante. Et dans la plupart des cas, si vous utilisez l'un d'entre eux, l'autre sera bloqué dans le même contexte ; vous devrez faire le choix.

Donc dans le cas du script latin, la plupart des règles des variantes sont des variantes bloquées. Dans les scripts, les variantes ont des membres pour le même script, mais il y a également des ensembles de variantes interscripts qui ont des membres de différents scripts, surtout dans le cas du latin, du cyrillique et du grec. Et il y a également des ensembles qui résultent de la de la combinaison des deux types. Diapositive suivante.

Chapitre 6. Ensemble avec les annexes D1 à D9, le chapitre 6 inclut des principes pour les ensembles de variantes et des données et des analyses des ensembles de variantes et des candidats à devenir des ensembles de variantes. Cela peut être étudié pour voir pourquoi nous avons tiré la conclusion qu'il s'agit de variantes.

Il y a deux ensembles de variantes qui sont spéciaux et ce, parce qu'ils sont liés à la version précédente de l'IDNA de 2003, mais également parce que cela inclut des règles qui permettent l'attribution d'autres variantes. Et ces deux ensembles correspondent au SS ou au S prononcé en allemand et au l avec et sans point. Ces deux ensembles sont spéciaux et sont particulièrement définis dans la proposition. Diapositive suivante.

Finalement, l'annexe E présente une liste des ensembles de variantes qui peuvent porter à confusion de par leur similarité visuelle. Il ne s'agit donc pas de variantes, nous ne les présentons pas comme telles,

elles ne font pas partie du fichier XML formel pour les règles de génération d'étiquette et l'annexe apparaît en tant que référence pour les personnes qui souhaitent analyser la similarité visuelle entre les deux chaînes. Merci.

Voici le processus. En ce moment, nous en sommes à l'étape de consultation publique et nous espérons pouvoir présenter la version finale d'ici janvier l'année prochaine. Nous vous prions donc de bien vouloir apporter des commentaires à notre proposition pour pouvoir y parvenir. Il reste à savoir bien sûr comment faire pour les questions. On verra s'il y en a, vous me le direz.

PITINAN KOOARMORNPATANA : Merci. Voyons s'il y a des questions dans le chat, ou si vous souhaitez prendre la parole, veuillez lever la main également pour nous le faire savoir. Je ne vois pas de demande d'intervention. S'il y a des questions pour le panel de génération en script latin, on reviendra vers vous.

En attendant, nous allons avancer et nous allons présenter les règles de génération d'étiquette en script japonais. C'est Hiro qui va les présenter.

HIRO HOTTA :

Merci Pitinan.

Bonjour à tous. J'espère que vous m'entendrez correctement.

zone racine

PITINAN KOOARMORNPATANA : Oui, tout à fait.

HIRO HOTTA :

Très bien. Ma connexion n'est pas très stable. Toutes mes excuses si nous avons des problèmes de connectivité à un moment ou à un autre, mais j'espère pouvoir me reconnecter tout de suite le cas échéant. Diapositive suivante et suivante.

Vous voyez ici les membres du panel de génération du script japonais. Et ici, la situation est similaire à celle des autres panels de génération, à savoir que nous avons des membres de différents secteurs. Diapositive suivante.

J'ai inclus ici un aperçu par rapport à la langue et au script japonais. Je pense que la langue japonaise présente un cas très particulier du fait que nous avons trois scripts. Nous écrivons même en ASCII et avec des numéros. Or, les trois scripts qui existent en japonais sont le kanji, le hiragana et le katakana. Ces caractères peuvent être mélangés dans n'importe quel ordre dans un mot et les règles de génération d'étiquette transmettent le JIS. On les définit en script ordinaire industriel japonais de premier et de deuxième niveaux, qui sont ce que l'on utilise au quotidien.

Nous avons plus de 6 000 caractères en japonais, donc beaucoup de caractères dans un répertoire. Et bien sûr, tout cela est défini dans notre MSR. Nous avons, je le disais à l'instant, trois types de scripts. Le kanji est commun aux langues du chinois et du coréen, ce qui présente des difficultés au moment de définir les variantes. J'en parlerai plus tard.

Outre le kanji, nous avons le hiragana et le katakana que nous avons dans notre pays. Mais dans le cas du hiragana, nous l'utilisons en tant que suffixe du kanji pour pouvoir compléter la lecture d'un mot. Pour les verbes, pour les conjonctions, nous écrivons les kanji sous différentes formes qui sont plus faciles à lire, par exemple pour que les enfants de cinq ans puissent les lire.

Le katakana est surtout utilisé pour d'autres types de mots. On l'utilise pour représenter de longs mots de langues étrangères, ainsi que des onomatopées.

Tous les caractères japonais sont considérés indépendants. En chinois et en coréen, il en est de même. Mais eux, ils ne partagent que certains caractères du kanji. Dans le cas de certains des caractères, on les considère des variantes lorsqu'il y a deux ou plus de caractères qui ont le même sens. Certaines personnes dans la communauté de la langue japonaise demandent que certains caractères soient considérés des variantes en chinois et en coréen également.

Nous avons ici des informations d'ordre général par rapport aux règles de génération d'étiquette pour la zone racine en japonais. Je vous disais que nous avons plus de 6 000 caractères, mais nous n'avons pas de variantes intrinsèques en japonais, ou pas de variantes qui dérivent d'un même sens ou d'une même prononciation.

Pour ce qui est des caractères qui sont identiques visuellement, nous avons défini 10 ensembles de variantes, ce qui est singulier puisque nous avons repris les variantes en kanji qui sont définies dans les règles de génération d'étiquette pour les scripts chinois et coréens. Je

pense qu'il y en a des centaines que nous avons importées de ces règles.

Et deuxièmement, autre point important, c'est que nous n'avons pas d'étiquettes pour les variantes outre l'étiquette originale qui puissent être attribuées. Et c'est cela la partie la plus difficile dans le cas de notre script japonais, mais nous avons finalement décidé de ne pas avoir d'étiquettes pour les variantes.

En japonais, nous suivons une règle spécifique très simple, et c'est que n'importe quel [inaudible], marque diacritique ou marque prolongée ne peut commencer une étiquette. Et cette règle s'applique à tous les mots ordinaires en japonais. Diapositive suivante.

J'aimerais attirer votre attention sur ce point, sur l'aspect identique au niveau visuel dans cette écriture japonaise. Premièrement, il pourrait y avoir un tiret et différents signes qui peuvent être des signes de ponctuation et qui peuvent prêter confusion par rapport à d'autres caractères. Nous avons donc là des variantes et nous devons choisir ce qui est identique visuellement et y travailler. Nous avons 10 paires qui ont été définies et choisies comme étant visuellement identiques. Nous avançons.

La confirmation de cet aspect d'être identiques visuellement dans ces 10 paires, basé sur l'Unicode, nous permet de voir si effectivement, cela prêtait à confusion si elles sont identiques. Nous avons eu 40 personnes qui ont examiné la situation, 20 personnes qui parlaient bien le japonais et 20 ans personnes faisant des tests ne parlant pas le japonais. Le résultat, c'est que toutes le paires ont un niveau de moins

de 3.2, donc cela veut dire qu'on peut en effet les confondre. Ce sont donc des variantes.

La deuxième recherche, c'était pour voir s'il y avait d'autres LGR que nous pouvions identifier. Nous avons fait une étude sur 176 divers r cipients. Nous avons eu un bon taux de r ponse avec 73 r ponses, et le r sultat, c'est que qu'aucune de ces pairs n'ont confondu plus de 3 % des r pondants, donc cela ne constituait pas des variantes.

Deuxi mement, en ce qui concerne les variantes que l'on peut affecter, toute combinaison de caract res est permise dans ces  tiquettes japonaises. Cela indique qu'il y a beaucoup de variantes et il y a  galement cette importation que nous avons vue du chinois et du cor en. Par exemple, vous avez des caract res kanji que vous voyez   l' cran o  il y a trois variantes. Et l'universit  de Keio utilise ces variantes dans le cadre de .jp. Comme vous le voyez, on peut utiliser ces variantes suivies de .jp. Si l'universit  de Keio a pu toutes les utiliser, ces quatre TLD sont permis dans la zone racine simultan ment. Mais une telle r gle pourrait faire un peu exploser la taille de la zone racine parce que nous aurions beaucoup de variantes. Nous aurions besoin de r duire le nombre d' tiquettes de variantes disponibles pour  viter cette explosion de la taille de la zone racine.

Nous avons d  avec ce groupe de g n ration japonais trouver des m thodes pour r duire le nombre d' tiquettes qu'on pouvait affecter. Nous devons trouver des m thodes pour r duire le nombre d' tiquettes et nous avons plusieurs m thodes. Nous avons finalement d cid  d'utiliser uniquement ces  tiquettes valides et de

bloquer toutes les étiquettes de variantes. Il y a ce blocage qui a été effectué.

Voilà comment nous avons décidé de progresser. Nous avons une période de commentaires publics qui se terminera le 16 novembre. Si vous avez des commentaires, n'hésitez absolument pas à nous faire savoir quels sont vos points de vue.

Merci beaucoup de votre attention.

PITINAN KOOARMORNPATANA : Merci Hotta-san. Je ne sais pas s'il y a des questions dans le chat ?

SARMAD HUSSAIN : Nous avons une question, mais elle n'est pas dirigée à la langue japonaise, elle est plus générale. Nous y répondrons possiblement à la suite de la présentation de Michel.

PITINAN KOOARMORNPATANA : Très bien.

Nous allons maintenant donner la parole à Michel Suignard, qui va nous parler de la version 5 de la LGR de la zone racine.

MICHEL SUIGNARD : Merci beaucoup.

La présentation va couvrir quatre thèmes. Tout d'abord, nous verrons l'intégration et la gestion pour les langues japonaises, chinoises et

pour le coréen, les étapes concernant la génération et le traitement des étiquettes. Ensuite, nous verrons les écritures latines, grecques, cyrillique et arméniennes et ensuite, la portée de cette version 5. Nous avançons.

Ça, c'est en rapport avec ce qu'Hotta-san nous a indiqué. Nous avons un travail à effectuer, c'est d'intégrer le chinois et le japonais et l'écriture coréenne. Nous avons terminé le travail sur la langue chinoise. Le seul ouvert pour des commentaires publics, c'est le japonais. Les autres langues, le chinois et le coréen, sont passées par cette période de commentaires publics.

C'était un travail assez important. Il y avait plus d'un millier de points de code et beaucoup de problèmes de variantes qui existaient. On pourra revenir là-dessus lors des questions. Nous avons travaillé à l'intégration de ces ensembles. Nous avons travaillé avec les différents groupes GP. Dans ces LGR, il y a des variantes qui ont été créées pour intégration. Et il y a deux ensembles. Il y a au moins deux points code résultants du LGR, donc intégration provenant du répertoire. C'était le cas pour le coréen et le japonais. Mais c'est le chinois qui a le plus de variantes en raison des versions traditionnelle et simplifiée de la langue chinoise. Beaucoup de caractères peuvent être simplifiés dans leur forme, donc version traditionnelle et version simplifiée, donc raison historique pour cette complexité. Il y a des caractères simplifiés qui existent également en japonais et en coréen et beaucoup de variantes au niveau de cette écriture traditionnelle. On en a parlé un petit peu plus tôt. Le LGR chinois a le plus de variantes parce que beaucoup de ces ensembles de variantes au moins deux membres. En

japonais et en coréen, la définition est un petit peu différente pour les LGR. Nous devons demander d'inclure pour les LGR japonais ou coréens ce qui a au moins deux points de code. C'était la première partie.

Ces ensembles de variantes étaient définis pour le japonais et pour le coréen. Et dans ces deux LGR, il y a eu un blocage qui a été effectué, mais il y a eu plus d'acceptation au niveau du LGR chinois, donc système un petit peu différent.

De plus, pour le LGR chinois, il y avait des variantes qui ont été soumises et il y avait eu une confusion entre le kana et le hangeul parce que le LGR japonais comme on l'a vu a créé des variantes entre les caractères kanji et les katakana. Ce n'était pas la même chose pour le chinois. Même chose pour le coréen. Le coréen est un petit peu différent – c'est pour le hangeul –, il y a des caractères un petit peu différents. Il y a des caractères chinois en coréen. Il y a une confusion entre le hangeul et des caractères chinois. Donc nous devons en prendre compte dans les règles de génération d'étiquette pour le chinois.

Nous avons vu ces deux points de code. Mais en fin de compte, nous avons dû inclure des variantes avec un seul point de code associé à des caractères chinois qui n'étaient pas de leur ensemble et qu'on va retrouver en coréen et en japonais.

Ceci dit, la collision des étiquettes sera déterminée dans le cadre du fichier fusionné LGR qui a l'ensemble complet. Donc c'est comme cela que nous avons travaillé avec ces étiquettes affectées uniquement

pour le chinois. Nous avons un LGR fusionné ou un LGR commun. C'est comme cela que l'on détecte les collisions possibles entre ces différents points. Nous avons des mécanismes pour déterminer les collisions possibles, donc un nouveau niveau qui est toujours disponible pour l'intégration.

Dans ces trois ensembles, les étiquettes que l'on pouvait affecter uniquement pour le LGR chinois ont des variantes. Mais pour clarifier les choses, il y a des variantes au niveau du japonais, mais elles sont bloquées comme cela a été expliqué. Nous ne pensons pas avoir de véritables problèmes. Nous attendons la période de commentaires publics. Voilà donc les différentes étapes pour la gestion d'une étiquette.

Vous avez un diagramme à l'écran que je voulais partager pour que vous voyiez un petit peu comment nous travaillons. Nous avons vraiment deux éléments lorsque nous soumettons une étiquette pour la zone racine. Il y a deux voies si vous voulez. Nous avons l'étiquette originale et il y a également l'étiquette variante. Vous avez ce qui est soumis pour les dossiers de demande et nous utilisons des éléments du LGR pour vérifier les variantes pour que cela satisfasse le répertoire et les règles qui existent pour cette génération d'étiquettes. Nous exprimons cela dans le LGR et nous déterminons cela de cette manière. Et ensuite, nous générons une variante d'index et c'est une représentation unique par étiquette qui peut être comparée avec les étiquettes existantes. Et si c'est unique, nous l'acceptons. Si ce n'est pas unique, nous le rejetons. Une autre variante de cette étiquette peut être déjà déléguée et dans ce cas, nous la rejetons.

La deuxième partie est un pas facultatif, à savoir qu'un candidat peut également proposer des étiquettes de variantes. Et cela est déterminé à travers ses propres recherches ou l'on cherche dans les règles de génération d'étiquette ce qui pourrait constituer une variante et que l'on pourrait demander. Puis, au moment de créer ces étiquettes de variantes, il faut les ensembles disponibles pour voir les éléments et savoir s'il s'agit de variantes qui sont susceptibles d'être attribuées. Dans la négative, cela sera bloqué et la candidature sera rejetée. Mais s'il s'agit d'une étiquette qui peut être attribuée, cela deviendra alors une variante attribuable qui pourra alors être déléguée dans la zone racine.

Il est important de signaler qu'il faut être prudent au moment de décider des étiquettes que l'on souhaite choisir, variante ou originale, parce que les étiquettes ne sont pas symétriques. Il y a une asymétrie entre les étiquettes des variantes et les étiquettes originales pour chacun des ensembles. Donc il faut faire attention à sélectionner l'étiquette originale qui permet l'existence également d'une variante parce que si vous sélectionnez directement l'étiquette de variante, la zone racine ne va pas la vouloir. Donc cette étiquette ne serait pas susceptible d'être attribuée parce qu'il n'y a pas de symétrie. Il faut faire attention à ce que vous sélectionnez en premier si vous voulez que les deux puissent être attribuées. Il faudrait voir les définitions des règles de génération d'étiquette et les ensembles de variantes utilisées et acceptées. Diapositive suivante.

Nous avons quatre scripts qui sont interconnectés, à savoir l'arménien, le cyrillique, le grec et le latin. La plupart d'entre eux sont

complétés, mais dans le cas du latin, on en est toujours à la consultation publique. Le script grec est déjà passé par la consultation publique il y a peu de temps. Dans le cas du cyrillique, cela fait longtemps, donc on peut déjà remettre les résultats de ce travail.

Une définition rapide dit que variantes comprises dans le système sont pleinement énumérées dans chacune des règles de génération d'étiquette. En définitive, les variantes vont dépendre de l'interdépendance de ces quatre scripts et cela apparaîtra dans chacune des règles de génération d'étiquette que vous pouvez consulter comme un ensemble cohérent. Et si vous trouvez une variante dans un LGR, vous pourrez l'utiliser en tant que telle.

Nous avons fait une exception dans ce que nous appelons des formes génériques, comme le haut par exemple qui est un cercle, ou le C. Il s'agit de formes qui existent également dans d'autres LGR qui n'ont rien à voir. Dans le script du grec, l'arménien, le cyrillique, le latin, vous trouverez des formes de scripts qui sont dans la zone racine mais qui ne sont pas liées à aucun de ces scripts. Nous avons décidé d'utiliser un mécanisme d'intégration qui énumère les LGR et ce sera intégré dans les LGR pour le latin lorsqu'elles seront publiées. On a également des règles qui ont été héritées des autres, mais nous avons supprimé cela afin d'éviter les confusions.

Comme je le disais avant encore une fois, il est nécessaire d'utiliser un fichier commun fusionné pour pouvoir déterminer la collision d'étiquettes et la validation est calculée toujours en fonction de ce fichier ; pas de changement là-dessus.

Nous avons une situation qui a été remise dans le cas de deux LGR, pour l'arménien et le cyrillique, parce que nous n'étions pas sûrs de la situation que cela donnerait pour les variantes des scripts. Lorsqu'un script partage des variantes avec un autre script, entre deux membres du même script, on pourrait avoir des problèmes. Il s'agit d'une situation assez fréquente, on l'a déjà eue dans le passé, mais les LGR latines ont imposé des variantes au cyrillique qui ne faisaient pas partie de la proposition du cyrillique à la base. Donc on a dû ajouter ces variantes dans le script et dans les LGR du script cyrillique.

Autrement dit, les LGR remises doivent incorporer ces propositions mises à jour si elles sont acceptables, mais on propose surtout d'avoir des schémas qui montrent les correspondances et que cela soit ajouté dans le cadre de l'intégration. Il a fallu que tout cela soit reconfirmé par les membres du GP pour être sûr que c'était acceptable. Dans la LGR 5 de la zone racine, vous pourrez voir la révision publique de l'ensemble pleinement intégré.

Alors en synthèse, nous avons 18 scripts existants dans le contenu du LGR 5 de la zone racine et nous avons complété ou ouvert deux consultations publiques, à savoir deux scripts pour chinois, japonais et coréen, donc le japonais et le coréen, et nous avons également incorporé deux nouveaux scripts alphabétiques, le latin et le grec. Nous avons deux scripts alphabétiques remis qui étaient le cyrillique et l'arménien et nous sommes en train de travailler sur un script supplémentaire qui est celui du Myanmar. Et nous espérons que ce script pourra être présenté pour consultation publique bientôt, je ne vais pas préciser de date ici. Bref, la règle de génération d'étiquette 5

de la zone racine devrait contenir 25 scripts, le thaana et le tibétain n'étant pas prêts pour l'instant. On parlait de 28 scripts avant. Pour moi, il n'y en avait que 27. Peut-être ai-je oublié un script.

PITINAN KOOARMORNPATANA : Oui Michel. Il y en a deux pour le japonais, katakana et hiragana dans le cas du japonais. Voilà pourquoi.

MICHEL SUIGNARD : Oui, d'accord, je vois le lien. Avant, je comptais pendant que je suivais la présentation et je n'étais pas sûr. Voilà. J'ai ici apporté quelques références dans la diapositive finale, mais pas la peine d'en parler et de le présenter. J'espère qu'on aura le temps pour les questions que vous pourriez avoir, si jamais.

PITINAN KOOARMORNPATANA : Merci.

Sarmad, vous voulez reprendre la question qui a été posée tout à l'heure ou y avez-vous déjà répondu sur le chat ?

SARMAD HUSSAIN : Merci Pitinan.

Je vais lire la question qui a été envoyée. Michel en a déjà un peu parlé, je pense qu'il l'a quelque part expliqué, mais on verra s'il a quelque chose d'autre à ajouter. Il s'agit d'une question de Donna Austin : « Est-ce que la conclusion qu'il n'y a pas de variantes dans le

contexte des LGR japonais a été également tirée dans d'autres scripts, les conclusion étant qu'il n'y a pas de variantes qui soient permises en japonais ? » Elle demande si c'est également le cas dans d'autres scripts.

MICHEL SUIGNARD :

Il y a des variantes, beaucoup des LGR sont des variantes, mais la plupart sont bloquées. L'attribution représente une difficulté. Il y a des variantes qui sont susceptibles d'être attribuées, mais du point de vue du DNS, cela n'est pas tout aussi simple. Nous avons travaillé pour veiller à ce que s'il y a un schéma qui puisse être défini, on minimise la quantité de variantes utilisées. Dans le cas du chinois, je pense qu'il y en a un maximum de cinq qui peuvent exister avec l'étiquette originale. Donc si vous avez cinq étiquettes régionales demandées, vous ne pouvez avoir que quatre variantes supplémentaires qui puissent être attribuées ; c'est le maximum.

Donc on a déjà des mesures d'atténuation en place pour les LGR dans ces cas-là. Mais dans la plupart des cas, on a pu trouver des solutions. Le cas du japonais était bien plus compliqué. On a beaucoup travaillé avec le GP du japonais pour trouver une solution qui permette de nous remettre à cette limite maximale d'étiquettes de variantes qui puissent être attribuées. On a beaucoup travaillé pour essayer de trouver une solution, mais nous n'avons pas été en mesure d'en trouver une parce qu'on aurait pu avoir une quantité infinie d'étiquettes. À ce moment-là, on aurait pu se trouver face à des étiquettes créées pour générer des interruptions. Même dans le cas des outils, on aurait pu se retrouver face à un manque de solution

possible absolue. Voilà pourquoi nous avons conclu finalement ensemble avec le panel de génération de japonais bien sûr que la seule solution possible était de bloquer les variantes. Mais les variantes existent, c'est juste qu'elles sont bloquées. Je précise : beaucoup de LGR vont bloquer les variantes, ce n'est pas spécifique ici au cas du japonais.

PITINAN KOOARMORNPATANA : Merci Michel.

Y a-t-il d'autres questions.

SARMAD HUSSAIN :

Oui, nous avons reçu une autre question de Robert Nkambwe qui demande : « Ce travail est-il un sous-ensemble de l'acceptation universelle et vice versa ? Excusez-moi si la question n'est pas intéressante, mais je viens d'arriver dans l'écosystème de l'ICANN. » On demande si ce travail représente un sous-ensemble de l'acceptation universelle et vice versa ; les deux sont liés. Il y a déjà eu quelques réponses sur le chat, mais Pitinan, si vous souhaitez répondre vous-même, ce serait bien aussi.

PITINAN KOOARMORNPATANA : Oui, bien sûr. Merci pour cette question Robert.

Ils sont liés. Dans un sens, les IDN représentent le mécanisme pour faire en sorte que les noms de domaine internationalisés puissent être enregistrés. Si le nom de domaine est en script local, toutes les

langues peuvent techniquement être utilisées. Mais le travail pour faire en sorte que la partie technique de l'attribution fonctionne et soit acceptée sur l'internet correspond aux IDN et à l'acceptation universelle, qui est une initiative spécifique.

Donc il n'y a pas que les IDN dont nous nous occupons ici, mais une partie de notre travail est de nous occuper que les IDN soient acceptés par les applications sur internet. Mais il y a également la question des nouveaux gTLD. Cela pourrait être de nouveaux gTLD longs ou courts, ce qui suit le point. En ce moment, on a les points plus longs pour *.photography* ou les plus courts comme *.run* en anglais. Donc l'acceptation universelle travaille tant aux IDN qu'aux noms de domaine non IDN. Voilà l'interconnexion entre les deux. L'acceptation universelle et les IDN travaillent main dans la main.

J'espère avoir répondu.

SARMAD HUSSAIN :

Oui.

Il y a encore un commentaire, mais je pense qu'il serait convenable que Michel rebondisse dessus, et Robert Nkambwe. Robert nous dit : « Je m'attendais à voir des tests rigoureux avant que cela ne soit intégré à la zone racine. Je n'ai pas vu cette étape dans le schéma que vous avez présenté. »

On a peu de temps Michel, mais peut-être que vous pourriez lui raconter quels sont les tests qui sont réalisés sur les propositions des LGR avant qu'elles ne soient intégrées dans la zone racine.

MICHEL SUIGNARD :

Oui. Nous avons demandé au panel de génération de présenter des étiquettes de test à chaque fois et avec cet ensemble d'étiquettes, nous décidons des règles de répertoire des LGR proposées. En même temps, nous vérifions que les TLD existants dans la zone racine puissent être validés par les LGR, donc on valide effectivement les étiquettes existantes dans la zone racine.

Nous testons par ailleurs les règles de génération d'étiquettes vis-à-vis des données, les étiquettes existantes, pour chaque script. Vous pourrez tout trouver en ligne. Et d'ailleurs, la plupart des mots que l'on utilise fréquemment sont validés par les LGR. Et puis, à chaque fois, lorsqu'on voit des anomalies, on consulte les panels de génération pour savoir quelle en est la raison. Parfois, il y a des révisions, il y a des raisons historiques également pour lesquelles on voit des effets inattendus, mais à chaque fois, nous devons vérifier et être satisfaits du résultat.

Nous avons différents mécanismes de tests à travers lesquels nous mettons à l'essai ces étiquettes, mais l'intégration est surtout mécanique. Il s'agit de pouvoir présenter les LGR sous une forme qui soit raisonnable. Mais la partie technique est complétée par les panels de génération en intégration avec les équipes informatiques bien avant cela.

PITINAN KOOARMORNPATANA : Merci Michel, bien répondu à la question.

Nous en sommes déjà à l'heure de conclusion de cette réunion que je vais tout de suite déclarer clôturée. Merci de nous avoir rejoints et d'avoir participé. Je vous invite à consulter les deux procédures de consultation publique. Nous espérons pouvoir recevoir vos retours là-dessus. Et sur ce, je vais conclure cet appel.

Ayez une bonne fin de journée et une bonne réunion ICANN. J'espère vous revoir la semaine prochaine. Au revoir.

Nous allons maintenant arrêter l'enregistrement.

[FIN DE LA TRANSCRIPTION]