

Aspects liés au nommage des messageries internationalisées

2 novembre 2020



Table des matières

Introduction	3
Résumé analytique	4
Aspects politiques à prendre en compte	5
Scripts pris en charge	5
Longueur d'une chaîne de nom de messagerie	5
Mélange de scripts	6
Points à prendre en compte eu égard aux scripts écrits de droite à gauche (RTL)	9
Points à prendre en compte en matière d'alias et de noms affichés	12
Signes et symboles	13
Normalisation de caractères Unicode	14
Points à prendre en compte en matière d'équivalence	14
Autres points à prendre en compte	17
Glossaire	17



Introduction

Le présent document passe en revue différents points liés au nommage de messageries internationalisées et est destiné aux administrateurs de systèmes de messagerie. Les normes relatives aux e-mails [RFC6533 Notifications de statut d'envoi internationalisé et d'élimination](#) et [RFC6430 Valeur de type de rapport de commentaires d'e-mail : non spam](#) permettent désormais aux noms de messagerie (la partie d'une adresse électronique avant le signe '@') et aux noms de domaine (la partie après le signe '@') d'utiliser presque tout terme ou système d'écriture ; ils ne sont pas limités aux lettres A-Z et aux chiffres 0-9.

Il s'agit d'une amélioration notable pour les utilisateurs qui utilisent et préfèrent les termes non anglais, et pour les personnes avec lesquelles ils échangent des messages. Toutefois, les administrateurs systèmes doivent adopter des politiques relatives aux noms de messagerie afin de maximiser les bénéfices liés à un plus grand choix de caractères autorisés, tout en limitant la confusion, les erreurs et les intentions malveillantes.

Le présent document est destiné aux administrateurs de systèmes de messagerie qui souhaitent mettre à disposition de leurs utilisateurs des messageries internationalisées et/ou configurer et gérer des systèmes compatibles avec les adresses électroniques internationalisées. Il aidera les administrateurs à poser les bonnes questions et à prendre les bonnes décisions lors de l'élaboration de leur politique.

Dans ce document, on suppose que le lecteur a une bonne connaissance et expérience des systèmes de messagerie et de l'étape 1 de l'internationalisation des adresses de courrier électronique (EAI), c'est-à-dire qu'il est capable d'envoyer des e-mails vers des adresses internationalisées et de recevoir des e-mails provenant de ces adresses. Ce document traite des problèmes liés à l'étape 2 de l'EAI, tels que le nommage et l'hébergement de messageries internationalisées.

La rubrique Glossaire fournit des définitions des termes spécialisés et acronymes utilisés dans le présent rapport.



Résumé analytique

Le présent document est destiné aux administrateurs de messageries électroniques, aux administrateurs systèmes et aux responsables informatiques. Il a pour but de regrouper et de préciser les principaux aspects opérationnels et politiques à prendre en compte lors de la création de noms de messageries électroniques internationalisées, d'alias et de noms affichés.

Les noms de messagerie (la partie de l'adresse électronique se trouvant avant le symbole '@') déterminent l'utilité, la facilité d'utilisation et la sécurité d'un système de messagerie électronique. Par conséquent, il est important que les administrateurs de messageries électroniques adoptent de bonnes politiques pour le nommage de ces messageries. Traditionnellement, les adresses électroniques étaient limitées aux lettres latines et aux chiffres, mais la technologie des adresses électroniques a évolué et permet désormais aux noms de messagerie et aux noms de domaine d'être écrits dans presque l'ensemble des langues et scripts. Cette souplesse complexifie les choix politiques, et le présent document vise à orienter les administrateurs de messageries électroniques dans l'élaboration d'une politique de nom de messagerie permettant de prendre en charge des adresses électroniques ne contenant ni lettres latines ni chiffres. Enfin, l'objectif commercial des systèmes de messagerie ainsi que les exigences linguistiques des utilisateurs et de leurs correspondants détermineront les choix politiques et les scripts (systèmes d'écriture pour les langues) autorisés dans les noms de messagerie et la longueur de ces derniers.

Étant donné que l'orthographe, certaines combinaisons de caractères, certains signes ou certains symboles peuvent prêter à confusion ou être trompeurs, la politique utilisée par les administrateurs doit, le cas échéant, les interdire (les détails de cette interdiction dépendront de la langue utilisée pour le nom de messagerie). Les problèmes techniques liés au fait que des noms peuvent sembler similaires pour des utilisateurs alors qu'ils contiennent des caractères différents requièrent une attention particulière. Les scripts écrits de droite à gauche (RTL) sont particulièrement à même de créer des noms dont l'orthographe prête à confusion, et donc si votre système autorise ces scripts, d'autres aspects politiques doivent être pris en compte.

Les utilisateurs peuvent échanger des e-mails avec des personnes ayant des difficultés à lire le script utilisé dans une adresse électronique. Afin de pallier ces difficultés, il est recommandé de mettre en œuvre une politique permettant aux utilisateurs de choisir des noms affichés (le nom personnel affiché) et des alias d'utilisateurs (autres adresses d'un même utilisateur).

Le présent document comprend également des ressources et des références fournissant de plus amples informations sur plusieurs de ces aspects politiques. Il comprend également un glossaire de termes, certains pouvant être nouveaux pour les administrateurs adoptant de multiples scripts et langues pour leurs noms de messagerie.



Aspects politiques à prendre en compte

Le présent document traite d'aspects susceptibles d'être inclus dans une politique de nommage de messagerie. Chaque rubrique traite aborde l'un de ces aspects de la façon suivante :

- Description de l'aspect en question.
- Explication des points à prendre en compte selon les systèmes de messagerie, la base d'utilisateurs et les besoins de la base d'utilisateurs.
- Recommandations politiques (dans certains cas).
- Liste de contrôle des mesures à prendre.

Scripts pris en charge

Lors de la mise en place d'un système de messagerie, les administrateurs de messageries doivent décider quels scripts seront les scripts par défaut du système.

Le terme « script » est utilisé dans son sens linguistique et désigne les conventions et ensembles de caractères utilisés afin d'écrire dans une ou plusieurs langues. Par exemple, le script latin est l'ensemble composé des lettres A-Z et d'autres caractères couramment utilisés afin d'écrire en anglais, en français et dans bien d'autres langues d'Europe, d'Amérique du Nord et du Sud, et d'Afrique. Le script arabe est utilisé afin d'écrire dans de nombreuses langues du Moyen-Orient, d'Afrique du Nord et d'Asie du Sud.

Points à prendre en compte :

- La plupart des systèmes incluront des points de code ASCII à des fins de rétrocompatibilité. L'EAI permet aux administrateurs d'utiliser des scripts supplémentaires.

Liste de contrôle

[] Déterminez les attentes de vos utilisateurs et le marché cible afin de comprendre les systèmes d'écriture requis pour le nom de messagerie et le nom de domaine d'une adresse électronique.

[] Lisez ce document afin de comprendre les complexités liées à la mise à disposition d'autres scripts (par exemple risque de confusion, d'erreur et d'intention malveillante) et tout impact potentiel sur vos utilisateurs.

[] Selon les attentes de vos utilisateurs, dressez une liste des scripts autorisés pour les noms de messagerie.

Longueur d'une chaîne de nom de messagerie

Réfléchissez à la longueur du nom de messagerie en caractères. Quelles sont les longueurs minimum et maximum que vous souhaitez autoriser ?

Points à prendre en compte :

- Des contraintes liées au système peuvent peser sur la chaîne de nom de messagerie. Bien que les normes relatives aux e-mails autorisent jusqu'à 64 octets dans un nom de messagerie, certains systèmes imposent à la chaîne une limite de caractères inférieure à la limite de caractères standard.



- Les noms courts peuvent être bénéfiques à l'utilisateur car ils peuvent être plus faciles à retenir et à écrire. Les noms longs peuvent aussi être bénéfiques à l'utilisateur car ils peuvent être plus spécifiques et moins susceptibles d'entrer en conflit avec d'autres noms de messagerie. Vous avez peut-être déjà une politique imposant aux noms de messagerie d'être basés sur le nom personnel de l'utilisateur, par exemple son prénom et nom de famille séparés par un « . ».
- Conservez votre politique existante mais pensez à l'étendre à une plus grande variété de caractères et langues. Aussi, déterminez les langues et caractères que vos utilisateurs et leurs correspondants pourraient être en mesure de comprendre, de retenir et de saisir dans les champs d'adresse électronique.
- Lors de l'utilisation d'UTF-8 à des fins d'ajout de langues et de scripts, un seul caractère peut être représenté en de multiples octets, ce qui augmentera le nombre de caractères du nom de messagerie. Par exemple, un nom de messagerie qui comprend des lettres ASCII requérant 20 octets peut être étendu à 60 octets pour la représentation de 20 lettres dans une autre langue ou un autre script.

Recommandations : Aucune.

Liste de contrôle :

[] Informez-vous sur les limites de votre système.

[] Soyez au fait des attentes de vos utilisateurs liées aux noms de messagerie internationalisés.

[] Pensez à appliquer aux noms de messagerie internationalisés la même politique ou une politique similaire en matière de longueur à celle appliquée aux noms de messagerie historiques.

Mélange de scripts

Le mélange de scripts fait référence à l'utilisation de caractères issus de plusieurs scripts dans un même nom de messagerie de telle sorte qu'il ne s'agisse pas d'une écriture normale. Plus précisément, certains caractères issus d'autres scripts peuvent être similaires au point d'être confondus avec les caractères issus du script principal, ce qui comporte des risques d'intention malveillante ou de confusion.

Points à prendre en compte :

- Le mélange de scripts entraîne des problèmes de sécurité et un risque de confusion chez l'utilisateur. Par exemple, l'image ci-dessous montre plusieurs points de code Unicode différents mais d'apparence similaire, entraînant un risque de confusion chez l'utilisateur¹.

¹ Source : <https://thenextweb.com/wp-content/blogs.dir/1/files/2014/08/confusables-3.png>



a a a a a a a a a a a
b Ъ Ъ Ъ Ъ b b b b b b
c c c c c c c c c c c
d d đ đ d d d d d d d
e e e e e e e e e e e

- Vos correspondants écriront le nom de messagerie et le nom de domaine ensemble en tant qu'adresse électronique. Si le nom de domaine s'écrit dans un script donné, il est normal que le nom de messagerie s'écrive dans le même script. Déterminez si d'autres scripts peuvent être utilisés en plus du script du nom de domaine.
- Certaines combinaisons peuvent s'avérer nécessaires, tandis que d'autres peuvent prêter à confusion ou être malveillantes.
- Certaines cultures utilisent de multiples scripts dans leur écriture de tous les jours. Par exemple, la plupart des textes japonais sont un mélange de caractères issus des scripts kanji, hiragana, katakana et latin (« romain »). Toutefois, il n'est pas inhabituel de voir du script arabe dans des textes japonais.
- Les noms de messagerie en script latin ne sont en général pas sensibles à la casse. Les lettres en majuscules et en minuscules sont traitées de manière identique. Pour les scripts que vous utilisez, déterminez s'il existe d'autres scripts équivalents que vos utilisateurs pourraient vouloir utiliser. Dans un souci de clarté, on ne considère en général pas les mélanges de majuscules et minuscules comme un mélange de scripts ; consultez la [rubrique Points à prendre en compte en matière d'équivalence pour de plus amples informations](#).
- Au sein du système des noms de domaine (DNS), les noms de domaine internationalisés (IDN) peuvent être représentés sous deux formats : l'étiquette U et l'étiquette A. Dans le présent document, l'utilisation de l'étiquette A n'est pas considérée comme un mélange de scripts ; consultez la [rubrique Points à prendre en compte en matière d'alias pour de plus amples informations](#).

Recommandations :

- Pour une utilisation optimale, il est recommandé que le nom de messagerie utilise le même script que la partie nom de domaine de l'adresse électronique, sauf indication contraire afin d'atteindre les objectifs de votre organisation.
- Évitez le mélange de langues utilisant le même script dans le nom de messagerie afin d'éviter d'utiliser ou de créer des caractères de nom de messagerie visuellement similaires qui utilisent différents points de code.
- Seuls les caractères qui sont utilisés conjointement dans l'écriture ordinaire d'une langue doivent être autorisés pour un nom de messagerie donné dans cette langue. Dans les noms de messagerie, le mélange de scripts doit être interdit. Par exemple, une messagerie en arabe pourrait être limitée aux seuls caractères arabes, alors qu'un nom de messagerie japonais pourrait être autorisé à inclure des caractères kanji, hiragana, katakana et latin.



Exemple 1 : Adresse électronique en script arabe : `مستخدم@رسيل.السعودية`

Exemple 2 : Adresse électronique en script japonais avec mélange de scripts :

`誰か@日本語ドメイン名試験.jp`

Liste de contrôle :

[] Déterminez le risque de confusion, d'erreur et d'intention malveillante dû au mélange de scripts à la fois dans le nom de messagerie et entre le nom de messagerie et le nom de domaine.

[] Autorisez un mélange de scripts limité en cas de nécessité claire de l'utilisateur fondée sur les pratiques locales et les exigences du marché. Par exemple, les utilisateurs d'un service de messagerie pour la vente au détail peuvent demander des noms avec des types de mélange de scripts, alors que les membres d'une organisation peuvent avoir moins d'objections à un choix limité de noms de messagerie excluant le mélange de scripts.

[] Si les utilisateurs sont autorisés à demander leurs noms de messagerie, donnez-leur des exemples.

Prévenir les chaînes invalides et les chaînes à l'interprétation instable

Dans certains systèmes d'écriture, les caractères peuvent être combinés selon des schémas valides permettant une communication productive, mais également selon des schémas invalides entraînant des risques de confusion et d'atteinte à la sécurité.

L'interprétation de telles chaînes de caractères ne sera pas uniforme entre les plates-formes et les applications ; elle sera différente selon les systèmes et les affichages. Cet aspect propose des moyens d'éviter des chaînes invalides.

Points à prendre en compte :

- Le DNS a défini des règles claires d'utilisation ou d'interdiction de caractères pour les langues et les scripts. Pour certains scripts, notamment les scripts d'Asie du Sud-Est et d'Asie du Sud, les caractères et signes peuvent être combinés à plusieurs niveaux.
- Par exemple, prenez les chaînes en script thaï « `ดี` » et « `ดี` ». La première, « `ดี` » U+0E14(๕) + U+0E35 (ี) + U+0E35 (ี) est valide car elle ressemble à `ดี` U+0E14(๕) + U+0E35 (ี). La communauté d'utilisateurs du script thaï a élaboré une règle selon laquelle U+0E35 ne peut suivre U+0E35.
- Ces règles utilisées dans le DNS sont connues sous le terme de « Règles de génération d'étiquettes » (LGR). Il se peut que vous souhaitiez adopter ces règles pour les messageries de votre système.
- Toutefois, les LGR peuvent être plus restrictives que nécessaire eu égard à un système donné. Par exemple, les LGR pour la zone racine interdisent l'utilisation de chiffres, alors que vous pouvez les utiliser dans vos noms de messagerie. Certains scripts n'ont pas encore de LGR ; vous ne pouvez donc pas les utiliser comme référence.
- Vous pouvez vous référer aux [LGR de référence pour le second niveau](#) existantes afin de valider les noms de messagerie, et un outil de validation des LGR est disponible en ligne à l'adresse suivante : <https://lgrtool.icann.org/>. Le code source



pour l'outil est également disponible sur GitHub ; vous pouvez le télécharger et l'intégrer à votre système.

Recommandations :

- Évitez les potentiels problèmes liés à la sécurité causés par la création de noms de messagerie utilisant des chaînes invalides et leurs équivalents.
- Si vous souhaitez proposer des noms de messagerie dans une langue, reportez-vous aux LGR pour cette langue. Par exemple, si vous voulez proposer des noms de messagerie en allemand, utilisez les LGR pour l'allemand.
- Si vous souhaitez proposer des noms de messagerie dans un script qui pourrait couvrir plusieurs langues, référez-vous aux LGR du script en question. Par exemple, si vous voulez proposer des noms de messagerie en allemand, en français et en suédois, utilisez le sous-ensemble correspondant des LGR pour le script latin.
- Pour les scripts non couverts par les [LGR de référence pour le second niveau](#), vous pouvez consulter les LGR pour la zone racine ou suivre les principes fondamentaux du [RFC6912](#). Pour les règles sur les chiffres dans les noms utilisant des scripts écrits de droite à gauche, consultez le [RFC5893 Scripts écrits de droite à gauche pour les noms de domaine internationalisés dans les applications \(IDNA\)](#).

Liste de contrôle :

Déterminez si les LGR de référence pour le second niveau satisfont vos exigences concernant les chaînes de nom de messagerie. Si c'est le cas, adoptez les LGR de référence. Si ce n'est pas tout à fait le cas, modifiez-les selon vos besoins.

Diffusez les Règles de génération d'étiquettes (LGR) en tant que mécanismes permettant à vos utilisateurs finaux de comprendre quels caractères et combinaisons seront jugés valides.

Le cas échéant, utilisez l'outil de validation des LGR afin de valider manuellement les chaînes de messagerie.

Déterminez si vous souhaitez utiliser l'outil de validation des LGR tel quel ou si vous souhaitez le déployer sur un serveur interne afin d'optimiser votre flux de travail et de conserver les noms de messagerie qu'il traite dans votre réseau.

Points à prendre en compte eu égard aux scripts écrits de droite à gauche (RTL)

L'arabe, l'hébreu, le persan et l'ourdou sont les systèmes d'écriture de droite à gauche les plus importants actuellement. Au fil du temps, avec la progression de l'utilisation du script arabe, le répertoire de 28 caractères utilisés afin d'écrire la langue arabe a été complété afin d'intégrer les sons d'un bon nombre d'autres langues telles que le pachtou, le persan, l'ourdou, etc. Alors que l'alphabet hébreu est utilisé afin d'écrire en hébreu, il est également utilisé afin d'écrire dans d'autres langues telles que le yiddish.

Avec un script écrit de droite à gauche (RTL) et de haut en bas, l'écriture commence à droite de la page et se poursuit vers la gauche. C'est différent des systèmes d'écriture de gauche à droite (LTR) dans lesquels l'écriture commence à gauche de la page et se poursuit vers la droite.

Toutefois, il convient de noter que pour un script écrit de droite à gauche, une partie du texte peut être écrite de gauche à droite. Par exemple, en arabe, les numéros et les termes en



script latin sont écrits de gauche à droite. Ainsi, ces scripts sont parfois qualifiés de « scripts bidirectionnels ».

Points à prendre en compte :

- À la différence des scripts écrits de gauche à droite, les scripts écrits de droite à gauche peuvent prêter à confusion. Les Règles de génération d'étiquettes (LGR) pour ces scripts tiennent dûment compte de ces risques. Prenez au sérieux les autorisations et interdictions prévues par les LGR car il s'agit de règles pertinentes même si leur raison d'être n'est pas évidente.
- Lors de l'échange d'e-mails entre une adresse écrite de droite à gauche et des personnes ne lisant pas de scripts écrits de droite à gauche, il est même encore plus dur de s'assurer que l'adresse a été comprise. Le recours à des alias peut s'avérer particulièrement utile. Consultez la rubrique [Points à prendre en compte en matière d'alias et de noms affichés pour de plus amples informations](#).
- Certains scripts écrits de droite à gauche (tels que l'arabe et l'hébreu) peuvent avoir les « mêmes » étiquettes avec plus d'une représentation. Ces alternatives peuvent être traitées comme des équivalents à une adresse électronique plutôt que des adresses indépendantes. Consultez la rubrique [Points à prendre en compte en matière d'équivalence pour de plus amples informations](#).
- Il n'est pas rare qu'un texte écrit de droite à gauche soit bidirectionnel. Par exemple, des termes du script latin (tels que « Cola ») et des chiffres arabes (tels que 123) seront écrits de gauche à droite. De même, des signes de ponctuation adoptent la direction du texte les accompagnant. Cela signifie qu'il est plus probable que les noms de messagerie qui ont des composants écrits de gauche à droite comportent des fautes de frappe.
- Le script arabe utilise ses propres chiffres en plus des chiffres ASCII. Si un nom de messagerie comprend un chiffre, déterminez le type de chiffre à utiliser et déterminez si le nom utilisant l'autre type de chiffre doit être traité comme un « équivalent ». Les craintes liées au mélange de scripts valent aussi pour les noms ayant les deux types de chiffres.
- Avec du texte bidirectionnel, l'ordre des caractères peut ne pas être très clair visuellement. En règle générale, le texte est saisi et stocké dans le même ordre que son écriture, mais il se peut qu'il n'ait pas de correspondance simple avec la position du texte écrit de droite à gauche dans l'adresse affichée. Cela ne pose généralement pas de problème pour une personne sachant lire le script et comprendre le changement de direction. Mais une personne ne sachant pas lire le script peut avoir du mal à déterminer l'ordre d'actionnement des touches à partir de l'adresse telle qu'affichée.
- Les claviers locaux sont adaptés au script local et peuvent ne pas prévoir tous les caractères utilisés dans les adresses électroniques. Certains claviers n'ont pas de touches ('.') ou ('@'). Il est ainsi plus difficile de saisir les adresses électroniques. Déterminez le type de clavier qui sera utilisé pour vos noms de messagerie et s'il convient ou non d'autoriser des caractères qui sont plus difficiles à saisir.
- De nombreux scripts écrits de droite à gauche comprennent des signes diacritiques pour la représentation des voyelles et la prononciation. Ils sont en grande partie facultatifs. Si un nom de messagerie autorise ces signes diacritiques, les utilisateurs peuvent oublier de les inclure. Il est plus simple d'autoriser des noms sans ces signes diacritiques.
- Le script arabe, qui affiche habituellement des caractères joints ensemble par calligraphie, dispose de caractères facultatifs afin de forcer ou d'empêcher la jonction des caractères. On parle de « liant sans chasse » ou d'« antiliant sans chasse ». Il s'agit d'un mécanisme moins connu, un artefact de texte codé, ce qui explique pourquoi de nombreux utilisateurs ne savent pas comment l'utiliser. Il peut être plus



difficile pour des utilisateurs de saisir correctement des noms de messagerie qui reposent sur des liants.

Recommandations :

- Il n'est pas recommandé d'utiliser des signes combinatoires, surtout avec le script arabe. Par exemple, utilisez محمد au lieu de مُحَمَّد et شؤون au lieu de شُؤون .
- Les noms de messagerie qui utilisent des scripts écrits de droite à gauche ne doivent pas commencer par un chiffre. Des chiffres peuvent être placés au début d'un texte en vertu des règles de direction du texte et prêter ainsi à confusion.
- Évitez les potentiels problèmes liés à la sécurité causés par des chaînes invalides et leurs équivalents. Les chaînes invalides et leurs équivalents sont particulièrement courants pour le script arabe. Consultez la [rubrique Prévenir les chaînes invalides et les chaînes à l'interprétation instable](#) ci-dessus et la [rubrique Points à prendre en compte en matière d'équivalence](#) ci-dessous.
- Si vous souhaitez proposer des noms de messagerie dans une langue, reportez-vous aux LGR pour cette langue. Par exemple, si vous voulez proposer des noms de messagerie en arabe, utilisez alors les LGR pour l'arabe. Consultez la [rubrique Points à prendre en compte en matière d'équivalence](#) pour de plus amples informations.
- Si vous souhaitez proposer des noms de messagerie dans un script qui pourrait couvrir plusieurs langues, référez-vous aux LGR du script en question. Par exemple, si vous voulez proposer des noms de messagerie en arabe, en pachtou et en ourdou, utilisez alors [les LGR pour le script arabe](#).
- Pour les scripts non couverts par les [LGR de référence pour le second niveau](#), vous pouvez consulter les LGR pour la zone racine ou suivre les principes fondamentaux du [RFC6912](#).
- Pour les règles relatives aux chiffres dans les noms de messagerie utilisant les scripts écrits de droite à gauche, consultez le [RFC5893](#) qui fournit les directives nécessaires pour l'utilisation de scripts écrits de droite à gauche pour les noms de domaine internationalisés dans les applications (IDNA).

Liste de contrôle :

[] Évitez le mélange de scripts pour les scripts écrits de droite à gauche car cela peut prêter à confusion et entraîner des problèmes liés à la sécurité.

[] Utilisez les LGR de référence pour le second niveau ou les LGR pour la zone racine.

[] Évitez d'autoriser deux messageries avec des noms de messagerie équivalents.



Points à prendre en compte en matière d'alias et de noms affichés

Certains systèmes de messagerie facilitent la possibilité pour les utilisateurs d'avoir plusieurs noms de messagerie ou plusieurs noms affichés envoyant des e-mails vers et recevant des e-mails de la même messagerie. Ces alias et noms affichés servent de nombreux objectifs et peuvent être utilisés dans plusieurs cas d'utilisation de l'EAI.

Voici quelques cas d'utilisation :

- Répondre aux demandes d'utilisateurs souhaitant des adresses électroniques avec plusieurs scripts, par exemple pour une carte de visite bilingue avec le script natif et le script latin.
- Trouver des solutions pour l'envoi et la réception d'e-mails avec des systèmes incompatibles.
- Faciliter la reconnaissance, par les expéditeurs, de l'adresse électronique et d'alias dans leur carnet d'adresses.

Points à prendre en compte :

- Évaluez les voies de transmission des e-mails de vos utilisateurs vers leurs correspondants et de leurs correspondants vers vos utilisateurs. Ayez bien conscience que les e-mails transmis dans ces deux directions peuvent être confrontés à différents systèmes, une direction pouvant donc bien marcher et l'autre poser des problèmes.
- Plus les correspondants sont variés et plus la gamme de voies de transmission d'e-mails depuis et vers votre système est large, plus vous aurez de risque de tomber sur un système de transmission ne parvenant pas à transmettre un e-mail vers ou depuis des adresses électroniques internationalisées.
- Autoriser les alias ou les noms affichés qui sont dans un script différent que le nom de messagerie pourrait engendrer des problèmes de sécurité. Toutefois, cela peut aussi être bénéfique lorsque l'adresse EAI est utilisée dans différentes communautés de langues.

Vous trouverez ci-dessous un exemple de nom affiché arabe pour une adresse électronique anglaise :

"مشرف@اختبار-سجل.مصر" <adminmail@datamail.asia>
Nom affiché Adresse électronique

Recommandations :

- Envisagez de fournir une adresse électronique ASCII en tant qu'alias pour vos adresses électroniques internationalisées. L'alias transmet des e-mails au même endroit que l'adresse internationalisée. L'utilisateur doit être en mesure de choisir d'envoyer un message ou une réponse depuis l'adresse internationalisée ou depuis l'alias e-mail ASCII.

Liste de contrôle :

[] Déterminez si votre système prend en charge des alias.

[] Si votre système prend en charge des alias, ajoutez une option 'création d'alias' à l'interface utilisateur lors du processus de sélection du nom de messagerie. L'utilisateur sera informé qu'il doit avoir un alias ASCII et sera autorisé à créer un alias associé à son nom de messagerie EAI.



[] Vous pouvez éventuellement ajouter une interface utilisateur autorisant l'utilisateur à ajouter des alias supplémentaires plus tard, si votre système prend en charge cette option.

[] Si vous attribuez des noms de messagerie à des utilisateurs au lieu de les autoriser à sélectionner eux-mêmes des messageries, attribuez l'alias ASCII en même temps que vous attribuez le nom de messagerie EAI.

Signes et symboles

Certains signes et symboles sont autorisés dans les noms de messagerie ASCII, par exemple le point (.), le tiret bas (_), le trait d'union (-) et le signe plus (+). Lorsqu'il sélectionne un nom de messagerie EAI, l'administrateur doit également déterminer les signes ou symboles Unicode dont l'utilisation doit être autorisée dans les scripts pris en charge.

Points à prendre en compte :

- Le principe d'inclusion (voir la [procédure LGR](#)) doit être pris en compte. Le principe d'inclusion invite à n'inclure que l'ensemble de caractères, par exemple les signes et symboles, qui sont requis par les utilisateurs, le point de départ étant une liste vide. Les symboles qui créent des problèmes de sécurité propres à votre mise en œuvre doivent être évités.
- N'autorisez que les signes et les symboles que vos utilisateurs et leurs correspondants peuvent lire et saisir. Le nom de messagerie doit pouvoir être saisi via des méthodes de saisie (IME) couramment utilisées.
- Définissez une politique, cohérente et compréhensible, prévoyant des limitations pour le nom de messagerie latin similaires à celles du nom de messagerie internationalisé.

Recommandations :

- Évitez des symboles mathématiques susceptibles d'engendrer des problèmes de sécurité tels que #, \$, %, =, ", *, < et >.
- Les émojis ne doivent pas être autorisés (voir le rapport consultatif du SSAC de l'ICANN [SAC95](#)).
- Les signes combinatoires qui ne peuvent être saisis doivent être interdits, par exemple « l'Arabic Hamza Above » (U+0654).

Liste de contrôle :

[] Par défaut, évitez d'utiliser les symboles mathématiques.

[] Par défaut, évitez d'utiliser des signes et symboles qui n'existent pas sur les claviers/dispositifs de saisie de votre utilisateur. Certains signes et symboles ne sont pas disponibles sur les dispositifs de saisie de certains marchés.

[] Soyez au fait des attentes de vos utilisateurs et du marché local. Si des signes et symboles sont requis pour votre marché, le point (.), le tiret bas (_), le trait d'union (-) et le signe plus (+) sont couramment utilisés et pourraient être ajoutés au répertoire.

[] Évaluez l'utilisation de signes supplémentaires (le cas échéant) et assurez-vous qu'elle n'engendre pas de problèmes de sécurité pour les utilisateurs finaux, notamment une confusion chez les clients, le croisement des scripts et autres atteintes à la sécurité.



Normalisation de caractères Unicode

La norme Unicode définit une procédure de normalisation de texte qui remplace les séquences équivalentes de caractères de sorte que deux textes équivalents seront réduits à la même séquence de points de code.

Unicode définit les formes normalisées de caractères qui peuvent être encodés de plusieurs façons étant donné qu'il existe différents types de formes normalisées telles que la forme de normalisation D (NFD) et la forme de normalisation C (NFC).

Exemple : Å peut être encodé par 212B ou 00C5. Ces deux formes font référence à la même messagerie et non pas à différentes messageries, et doivent être normalisées avec les mêmes NFD et NFC.

Il convient de noter que ces exemples utilisent des caractères latins mais qu'il existe des exemples similaires dans d'autres scripts.

De plus amples informations sur les autres types de normalisation, accompagnées d'exemples, sont disponibles sur le [site web du consortium Unicode](#).

Recommandations

- Les noms de messagerie doivent utiliser la NFC à des fins d'uniformité des noms de domaine, et aussi parce que les chaînes NFC, dans la plupart des cas, sont plus courtes que les chaînes NFD.
- En règle générale, les logiciels standard encodent déjà la normalisation Unicode dans leurs bibliothèques. Les administrateurs de messageries doivent connaître le type de normalisation du système qu'ils exploitent afin de pouvoir attribuer et gérer leurs messageries.
- Les noms de domaine utilisent la forme de normalisation C (NFC). Lors de la construction d'un système, veillez à ce que les bibliothèques prennent en charge la NFC à des fins d'uniformité.

Liste de contrôle

[] Déterminez le type de normalisation de votre système de messagerie.

[] Assurez-vous que votre logiciel de messagerie effectue des comparaisons de noms indépendantes des formes normalisées.

[] S'il est possible de sélectionner un type de normalisation, sélectionnez la NFC à moins que vous n'ayez de bonnes raisons d'utiliser d'autres formes.

Points à prendre en compte en matière d'équivalence

Lorsque des étiquettes de nom de domaine sont qualifiées de « similaires » ou d'« interchangeables » par la communauté, on parle de « variantes » d'étiquettes de nom de domaine. Lorsque des chaînes de messagerie électronique sont qualifiées de « similaires » par la communauté, le présent document parle de noms de messagerie « équivalents ».

Parfois, les noms équivalents sont utiles et peuvent être utilisés comme alias pour le nom de messagerie principal. Dans d'autres cas, ils prêtent à confusion et peuvent entraîner des problèmes de sécurité s'ils ne sont pas identifiés ou attribués à un autre utilisateur.



Points à prendre en compte :

- Dans les adresses électroniques ASCII, certains services de messagerie peuvent qualifier d'équivalents les lettres majuscules et les lettres minuscules.
- La communauté du DNS a beaucoup travaillé afin de définir des Règles de génération d'étiquettes (LGR) claires, comprenant l'identification de variantes d'étiquettes dans différents scripts. Consultez les LGR afin de déterminer si vous souhaitez les utiliser comme fondement de votre définition de noms de messagerie équivalents.

Exemples :

1. Équivalents chinois traditionnels et chinois simplifiés :

名称 qui signifie « nom » (U+540D U+79F0)

名稱 qui signifie « nom » (U+540D U+7A31)

名称 (Míngchēng, qui signifie « nom ») est utilisé en Chine continentale et 名稱 (Míngchēng, qui signifie « nom ») est utilisé à Taïwan et à Hong Kong. Toutefois, l'ensemble de la communauté chinoise pourrait reconnaître que ces deux termes sont les mêmes.

2. Équivalents en script arabe (le terme signifie « réseau » en arabe) :

شبكة (Unicode U+06C3 U+06A9 U+0628 U+0634)

شبكة (Unicode U+0629 U+06A9 U+0628 U+0634)

شبكة (Unicode U+0629 U+0643 U+0628 U+0634)

3. Caractères de pleine chasse et caractères de demi-chasse en scripts japonais et chinois. « J I M » vs « JIM », « 1 2 3 » vs « 123 », « アイオ » vs « アイオ ».

4. Équivalents en terme de casse :

John.Callen

john.callen

John.callen

john.Callen

5. Variante de nom et séparateur (peut aussi être qualifiée d'équivalent) :

john.callen

johncallen

- Pour un système de messagerie, vous pouvez envisager de transmettre tout e-mail adressé aux noms de messagerie équivalents à la même messagerie. Cela permet d'éviter la confusion du client, le croisement de scripts et autres problèmes de sécurité.
- D'autres équivalents peuvent être attribués à la même messagerie en tant qu'alias, mais ils peuvent alors devenir très nombreux (des centaines ou des milliers) et constituer un véritable défi en termes de gestion. Tâchez donc de trouver le bon compromis entre l'utilité des équivalents et leur trop grand nombre.
- Il convient de noter que les variantes dans la partie nom de domaine de l'adresse électronique peuvent avoir un impact similaire sur l'utilisateur. Le débat sur les noms de domaine et leurs variantes ne relève pas du présent document.



- Soyez conscients des éventuels types d'erreurs ou du risque de confusion lorsque des utilisateurs saisissent des adresses électroniques pour votre système. Le présent document traite de la confusion dans la partie nom de messagerie de l'adresse.
- Il existe aussi un risque de confusion eu égard à la bonne orthographe ou la normalisation de la partie nom de domaine de l'adresse. Les variantes de noms de domaine ne relèvent pas du présent document.

Recommandations :

- Après l'attribution d'un nom de messagerie, l'équivalent de ce nom de messagerie doit être identifié, apparaître comme enregistré et aucun autre utilisateur ne doit pouvoir l'enregistrer.
- Pensez à traiter les équivalents les plus utiles d'un nom de messagerie comme des alias (voir la [rubrique Points à prendre en compte en matière d'alias](#) du présent document), et envoyez les e-mails adressés à la messagerie équivalente au même endroit que le nom de messagerie d'origine. Toutefois, la gestion des équivalents peut s'avérer difficile, et tâchez donc de trouver le bon compromis entre l'utilité des équivalents et le trop grand nombre d'équivalents à traiter, au cas par cas.
- Passez en revue les règles pratiques d'équivalence selon la langue pour une messagerie équivalente. Par exemple, avec des noms de messagerie ASCII, il est pratique de traiter toutes les variations majuscules ou minuscules comme des équivalents. Avec un nom de messagerie en chinois simplifié comme 中国 qui veut dire « Chine », envisagez de traiter le nom en chinois traditionnel ayant le même sens, 中國, en tant qu'équivalent.
- Toutes les permutations de casse doivent être considérées comme la même messagerie.
- Lorsque plus d'un ensemble de chiffres est utilisé dans le même répertoire (ensemble de points de code possibles utilisés pour créer des noms de messagerie), ces chiffres doivent être qualifiés de points de code équivalents. Par exemple, un administrateur de messagerie en langue hindi peut considérer abc123 et abc१२३ comme des noms de messagerie équivalents :
Nom de messagerie : abc123
Nom de messagerie équivalent : abc१२३

Liste de contrôle :

[] Définissez une politique à même de déterminer les noms de messagerie « identiques » ou équivalents qui s'appliqueront à votre système en fonction du système d'écriture, des attentes des utilisateurs et des capacités techniques de votre mise en œuvre. Votre politique devra examiner les LGR, la casse, les séparateurs, les nombres et les symboles traités dans les rubriques Points à prendre en compte ci-dessus.

[] Déterminez si les [LGR de référence pour le second niveau](#) satisfont vos exigences en matière d'équivalence de chaînes de noms de messagerie. Si elles ne satisfont pas vos exigences, adaptez-les en tant que de besoin.

[] Partagez votre politique afin de permettre aux utilisateurs finaux de comprendre quels caractères et combinaisons seront jugés valides et quels caractères et combinaisons pourraient avoir des équivalents.



[] Lors de l'utilisation de chaînes équivalentes, déterminez si vous souhaitez utiliser l'outil de validation tel quel ou si vous souhaitez le déployer sur un serveur interne afin d'optimiser votre flux de travail et de conserver les noms de messagerie qu'il traite dans votre réseau.

Autres points à prendre en compte

La partie nom de domaine internationalisé d'une adresse électronique peut s'écrire avec des caractères Unicode ou avec un nom équivalent uniquement sous format ASCII commençant par « xn-- ».

Points à prendre en compte :

- La partie « xn-- » d'un nom de domaine est un détail technique fondamental. Idéalement, il n'est jamais nécessaire que l'utilisateur le voie. Toutefois, en période de transition au cours de laquelle tous les systèmes de messagerie ne prennent pas en charge les adresses électroniques internationalisées, la partie « xn-- » peut être une solution utile.

Recommandations :

- Écrivez les noms de domaine avec leurs noms internationalisés non ASCII. Évitez d'afficher le nom alternatif « xn-- ».
- Certains clients de messagerie pourraient ne pas lier automatiquement les formes d'étiquette U et d'étiquette A des noms de messagerie d'adresse électronique ; assurez-vous donc de la correspondance des deux étiquettes.

Glossaire

Alias : Adresse électronique transmettant des e-mails au même endroit en tant qu'adresse électronique alternative.

En règle générale, un alias est une adresse électronique ASCII qui fournit une adresse de secours pour une adresse

EAI dans l'hypothèse où un correspondant ne pourrait envoyer d'e-mails à l'adresse EAI du fait que des systèmes de transmission ne prennent pas en charge l'EAI.

ASCII : Désigne un ensemble de caractères limités constitué de lettres du script latin sans accents,

de caractères numériques et de quelques signes de ponctuation. Le terme ASCII est un acronyme de l'industrie informatique.

Nom affiché : Nom figurant à côté de l'adresse électronique.

Nom de domaine : Séquence de noms utilisée afin d'identifier un site web ou un serveur de messagerie, par exemple

« uasg.tech » ou « icann.org ».

EAI : Internationalisation des adresses de courrier électronique.

Internationalisation des adresses de courrier électronique : Pratique d'adresses électroniques comprenant l'ensemble



des langues et scripts pratiques, pas seulement ASCII ou le script latin.

Équivalent : Lorsqu'au moins deux points de code Unicode ont une signification commune ou sont qualifiés, d'un point de vue linguistique, d'équivalents par les autorités locales, débouchant sur des « variantes de caractères ». Par exemple, cela se produit entre des scripts chinois simplifiés et traditionnels, et cela peut prêter à confusion pour les personnes utilisant ces scripts.

IDN : Nom de domaine internationalisé.

IME (méthode de saisie) : Application qui permet à un clavier standard (tel qu'un clavier US-101) d'être utilisé afin de saisir des caractères et des symboles qui ne sont pas directement représentés sur le clavier.

Nom de domaine internationalisé : Nom de domaine ayant des caractères non ASCII.

Règles de génération d'étiquettes : Règles pour l'utilisation des caractères dans les noms de domaine.

LGR : Règles de génération d'étiquettes.

Nom de messagerie : Partie d'une adresse électronique identifiant l'emplacement de la messagerie sous un nom de domaine. D'autres termes sont également utilisés pour désigner la même chose, par exemple « nom de compte », « nom d'utilisateur », « partie UTF-8 » et « partie locale ».

Punycode : Algorithme permettant d'encoder Unicode dans le sous-ensemble de caractères ASCII limité utilisé pour les noms d'hôte Internet.

Étape 1 de l'EAI : Système de messagerie pouvant envoyer des e-mails à et recevoir des e-mails depuis des adresses électroniques internationalisées, mais ne pouvant pas héberger de messageries avec des noms internationalisés.

Étape 2 de l'EAI : Système de messagerie pouvant héberger des noms de messagerie internationalisés et envoyer des e-mails à et recevoir des e-mails depuis des adresses électroniques internationalisées.

Interprétation : Processus via lequel un système d'affichage informatique transforme une séquence de caractères en un affichage visuel sur l'écran.

Script : Conventions et ensembles de caractères utilisés ensemble afin d'écrire dans une ou plusieurs langues.

Script.txt : Fichiers de données associés à la base de données de caractères Unicode, identifiant les valeurs de propriété de script de tous les points de code (<http://www.unicode.org/Public/12.1.0/ucd/Scripts.txt>).

Unicode : Norme d'encodage de caractères regroupant la majorité des scripts et des langues écrites du monde. Voir <http://unicode.org>.



Variantes de noms : Noms qualifiés d' « identiques » par la communauté.