

SYSTEME DE SIGNALEMENT DE
PROBLEMES LIES A L'EXACTITUDE
DU WHOIS

RAPPORT PILOTE

PRÉSENTÉ À :
l'ICANN

PUBLIÉ PAR :
NORC à
l'Université de Chicago.

23 DECEMBRE 2014



at the UNIVERSITY *of* CHICAGO

Sommaire

Résumé exécutif	3
Conception de l'échantillonnage	5
Résultats	6
Conséquences de l'étude pilote	7
Introduction	9
Conception de l'étude pilote de l'ARS	11
Exigences du rapport de l'ARS	11
Étude pilote de l'ARS	12
Tailles des échantillons des gTLD	14
<i>Les pays du cadre d'échantillonnage</i>	<i>14</i>
Définitions d'exactitude	15
Approches automatiques et manuelles.....	17
Approche et méthodes de l'étude pilote de l'ARS	19
Points de vue de la validation	19
Catégories de l'échelle de notation.....	20
Validation de l'adresse électronique	20
Validation du numéro de téléphone	21
Validation de l'adresse postale	22
Conception de l'échantillonnage	22
Principales conclusions.....	26
Exactitude du point de vue de la validation	27
Exactitude par les gTLD	29
Exactitude par le bureau d'enregistrement	31
Conséquences des résultats	33
Examen des définitions.....	33
Déterminer un pointage pour l'exactitude	34
Taille de l'échantillon et niveaux de confiance.....	36
Pas de rapport sur la conformité.....	36

Explorer la validation de l'identité	37
PROCHAINES ETAPES.....	38
ANNEXE A. Taille des échantillons par gTLD et par pays.	39
Annexe B : Prochaines étapes pour l'élaboration du système de signalement de problèmes liés à l'exactitude du WHOIS (ARS).....	49
Commentaires et consultation publics à Singapour.....	49
Processus de correction des enregistrements inexacts : groupe consultatif pour la mise en œuvre de l'ARS	50
Annexe C : DESCRIPTIONS DÉTAILLÉES DES PROCESSUS DES FOURNISSEURS	52
Validation du téléphone : DigiCert	54
<i>Validation syntaxique du téléphone.....</i>	<i>54</i>
<i>Validation opérationnelle du téléphone</i>	<i>57</i>
Validation du courrier électronique : Strikelron.....	58
<i>Validation syntaxique du courrier électronique.....</i>	<i>59</i>
<i>Validation opérationnelle du courrier électronique</i>	<i>62</i>
Validation postale : Union postale universelle (UPU)	65
<i>Validation syntaxique de l'adresse postale</i>	<i>66</i>
<i>Validation postale opérationnelle</i>	<i>71</i>

Résumé exécutif

Le 8 novembre 2012, suite aux [recommandations de l'équipe de révision du WHOIS](#) créée en vertu de l'[affirmation d'engagements \(AoC\)](#), le [Conseil d'administration de l'ICANN a approuvé](#) une série d'améliorations sur la manière dont l'ICANN gère sa surveillance du programme WHOIS. Dans le cadre de ces améliorations, [l'ICANN s'est engagée](#) à identifier de manière proactive des inexactitudes dans les données de contact des gTLD contenues dans la base WHOIS et à renvoyer ces fichiers potentiellement inexacts aux bureaux d'enregistrement afin qu'ils en assurent le suivi. Pour mener à bien ce travail, l'ICANN a entamé la mise au point du système de signalement de problèmes liés à l'exactitude du WHOIS (ARS) : un processus destiné à mettre en place des évaluations régulières de l'exactitude du WHOIS et à en publier les résultats.

L'étude pilote, conçue pour isoler et évaluer des composantes clés du système, servira de base à l'ARS. Les résultats de l'étude pilote serviront à orienter le travail de conception de l'ARS.

L'étude pilote a été un effort de collaboration internationale dirigé par l'ICANN, avec le soutien du [NORC à l'Université de Chicago](#), [DigiCert, Inc.](#), [StrikeIron, Inc.](#), [l'Union postale universelle \(UPU\)](#), et [Whibse, Inc.](#) [Suite à un appel à propositions](#), la planification initiale de l'étude pilote a commencé le 27 août 2014, et la collecte de données a débuté le 4 septembre 2014. La validation de l'exactitude des enregistrements WHOIS sélectionnés pour révision a été achevée pendant la courte période comprise entre le 12 septembre et le 19 septembre 2014. L'analyse des résultats de validation a été conclue le 8 octobre 2014.

Dans l'étude pilote, à partir d'un échantillon des enregistrements WHOIS, ces organisations ont validé, respectivement, les champs téléphone, adresse de courrier électronique et adresse postale. La validation a été complétée par l'analyse des différents facteurs de l'exactitude syntaxique et opérationnelle des échantillons des enregistrements WHOIS qui sont expliquées ci-dessous.

Point de vue syntaxique : le point de vue syntaxique examine les données WHOIS dans le but de déterminer quelles sont les données existantes ou manquantes dans les champs de données WHOIS et évalue si ces données répondent aux normes spécifiques de formatage et de structure.

Point de vue opérationnel : le point de vue opérationnel examine les données WHOIS dans le but de déterminer si les informations de contact peuvent être utilisées dans la pratique pour réaliser la fonction prévue du type de données et consiste à recueillir et utiliser des données externes à l'enregistrement WHOIS pour informer une évaluation de l'exactitude.

Une évaluation de l'exactitude des données WHOIS, guidée par l'une des perspectives devrait se traduire dans les données étant classées dans une échelle de notation basée sur le degré auquel elles peuvent être considérées comme exactes. Les catégories de l'échelle d'évaluation établies pour les études sur l'exactitude du WHOIS décrivent l'exactitude dans un intervalle inclusif entre « pas d'erreur » le plus haut degré d'exactitude, et « erreur totale », qui représente le plus faible degré d'exactitude. Pour l'étude pilote, les dossiers ont été regroupés dans deux grandes catégories « exact » ou « inexact » qui englobent les échelles de notation ARS comme suit :

Catégorie « exacte »

- Pas d'erreur : indique que les données fournies, analysées à partir des trois points de vue pour évaluer l'exactitude, contiennent les informations importantes et qu'elles sont exactes ;
- Erreur minimale : indique que les données contiennent les informations importantes. Il est possible que des clarifications s'avèrent nécessaires ou bien que des informations supplémentaires soient demandées, mais les données fournies sont exactes ; et
- Erreur limitée : indique que les données ne contiennent pas au moins une partie des informations importantes, mais que, dans une certaine mesure, les données existantes sont jugées utiles.

Catégorie « inexacte »

- Erreur majeure : indique que les données ne contiennent pas les renseignements importants, et que, pour la plupart, elles sont inexactes.
- Erreur totale : indique que les données ne contiennent pas les renseignements importants ou que les données fournies sont majoritairement inexactes.

Dans l'étude pilote, les contractants de l'ICANN ont appliqué des critères de validation prédéterminés au courrier électronique, au téléphone et aux données postales des dossiers

WHOIS pour évaluer l'exactitude du point de vue syntaxique et opérationnel. Les résultats de validation ont été compilés dans des ensembles de données, séparés par type de données et par point de vue, et ces ensembles de données ont ensuite été analysés pour produire des statistiques exactes (pour les nouveaux gTLD et les gTLD plus anciens) stratifiées par gTLD, région géographique, et bureau d'enregistrement.

Conception de l'échantillonnage

L'étude pilote a débuté avec la sélection d'un échantillon de 100 000 enregistrements des fichiers de zone des gTLD. Étant donné que les nouveaux gTLD sont l'objectif principal de l'étude, un pourcentage élevé de ces domaines ont été échantillonnés, ce qui représente 25 pour cent des domaines représentés dans l'échantillon (au moment de l'échantillonnage, seulement 1,4 pour cent des domaines étaient des nouveaux gTLD). Afin de fournir des informations sur tous les gTLD, les petits gTLD ont été sur-échantillonnés. Plus précisément, tous les nouveaux gTLD avaient au moins 10 enregistrements sélectionnés (ou tous si le nombre de dossiers disponibles était inférieur à 10) et tous les autres gTLD (« plus anciens ») ont eu au moins 30 enregistrements sélectionnés (ou tous si le nombre de dossiers disponibles était inférieur à 30).

Une fois que les dossiers ont été sélectionnés pour l'échantillon, ils ont ensuite été interrogés par les recherches WHOIS afin de produire des données pour l'étude pilote. Les données brutes du WHOIS ont été analysées pour que les informations de l'enregistrement puissent être systématiquement interprétées lors de l'étape de validation de l'étude. Après l'analyse, tous les enregistrements d'adresses électroniques ont passé par les étapes de validation syntaxique et opérationnelle. Le nombre de requêtes réussies (et les enregistrements analysés) était au nombre de 98 821. Un sous-échantillon de 10 000 enregistrements a été sélectionné pour la validation syntaxique des adresses postales et des numéros de téléphone. Afin de fournir des estimations au niveau de la région géographique, NORC à l'Université de Chicago a déterminé le pays (et la région géographique) pour autant de noms de domaine que possible, et a sur-échantillonné les deux régions les plus petites : Afrique et Asie/Australie/Pacifique.

Un petit sous-échantillon de 1000 enregistrements, (encore une fois, les régions Afrique et Asie/Australie/Pacifique ont été sur-échantillonnées) a été considéré pour la validation opérationnelle des adresses postales et des numéros de téléphone. Le même ensemble de

variables a été utilisé pour classer les 10 000 enregistrements avant de sélectionner un échantillon systématique de 1000 enregistrements.

Un total de 200 pays a été représenté dans l'échantillon complet de 98 821 noms de domaine. Parmi ces 200 pays, 91 étaient représentés dans les sous-échantillons de 10 000 et de 1000 ainsi que dans l'échantillon complet (les pays d'Afrique et d'Amérique latine/Caraïbes ont été sur-échantillonnés dans les sous-échantillons).

L'échantillon complet a été utilisé pour la validation du courrier électronique. La vérification syntaxique pour les numéros de téléphone et les adresses postales a utilisé le sous-échantillon de 10 000, tandis que la vérification opérationnelle pour les numéros de téléphone et les adresses postales a utilisé le sous-échantillon de 1000. Les rapports d'analyse sur les sous-groupes avec au moins 20 enregistrements, et seuls les meilleurs sept gTLD ont répondu à ce critère dans le sous-échantillon de 1000.

Il est important de noter que les différentes tailles d'échantillons se traduisent par des déterminations très différentes dans les estimations ci-dessous. Plus précisément, pour la vérification du courrier électronique, de petites différences sont statistiquement significatives tandis que pour la validation opérationnelle des numéros de téléphone et des adresses postales, les différences de pourcentage pourraient ne pas être statistiquement significatives.

Résultats

L'exactitude de l'information du courrier électronique, de l'adresse postale et du téléphone a été évaluée du point de vue syntaxique et opérationnel, et les statistiques du groupe ont été calculées pour tous les gTLD, les RAA 2009 et 2013, pour les gTLD plus anciens et pour les nouveaux gTLD. Quelques résultats notables sont les suivants :

- les différences de l'exactitude syntaxique du courrier électronique et du téléphone ne sont pas statistiquement significatives pour les gTLD plus anciens et pour les nouveaux gTLD alors que l'exactitude syntaxique des adresses postales est meilleure dans les gTLD plus anciens.

- les chiffres opérationnels pour les adresses de courrier électronique indiquent que l'exactitude des nouveaux gTLD est légèrement supérieure à celle des gTLD plus anciens. L'exactitude opérationnelle des numéros de téléphone associés aux gTLD plus anciens est plus importante que celle des nouveaux gTLD mais les deux groupes affichent des résultats similaires pour l'exactitude opérationnelle des adresses postales.
- les différences dans l'exactitude syntaxique entre les bureaux d'enregistrement régis par le RAA 2013 et le RAA 2009 pour le courrier électronique, numéros de téléphone et adresses postales ne sont pas statistiquement significatives.
- du point de vue opérationnel, les bureaux d'enregistrement régis par le RAA 2013 affichent un taux d'exactitude des adresses de courrier électronique plus élevé que les bureaux d'enregistrement régis par le RAA 2009. Les bureaux d'enregistrement régis par le RAA 2009 ont plus d'exactitude opérationnelle pour les numéros de téléphone tandis que les deux groupes sont similaires quant aux adresses postales.

Conséquences de l'étude pilote

L'étude pilote a été un exercice très instructif, et a fourni de précieuses leçons et mises en garde pour les méthodes et procédures de l'ARS. Voici les principales leçons tirées du projet pilote :

- **processus lent** : la sélection des échantillons des enregistrements WHOIS qui transforment les informations WHOIS brutes en champs susceptibles de faire l'objet d'une évaluation, l'évaluation des adresses postales, des adresses électroniques et des numéros de téléphone prend du temps, et ne doit pas être précipitée. Le calendrier de l'étude pilote a été fortement condensé et devrait être allongé lorsque le système de rapport de l'exactitude du WHOIS est lancé.
- **validation de l'adresse postale** : contrairement à l'adresse de courrier électronique et au numéro de téléphone, une adresse postale est composée d'un certain nombre de champs. Cela fait qu'il soit plus difficile de valider les adresses postales et que les résultats soient différents de ceux du courrier électronique et du téléphone. L'échelle d'évaluation des cinq catégories (pas d'erreur, erreur minimale, erreur limitée, erreur majeure, erreur

totale) n'est pas nécessairement utile pour la validation de l'adresse postale, bien qu'elle puisse être utile pour la validation du courrier électronique et du téléphone.

- **questions liées à la taille de l'échantillon** : la validation des adresses postales et des numéros de téléphone prend plus de temps que celle des adresses de courrier électronique. En outre, la validation opérationnelle prend plus de temps que la validation syntaxique des adresses postales et des numéros de téléphone. Voilà la raison pour laquelle des échantillons de différentes tailles ont été utilisés. Cela signifie que les mesures de l'exactitude effectuées sur la base d'échantillons plus petits sont moins fiables que celles effectuées sur des échantillons plus grands. En particulier, la validation opérationnelle pour l'adresse postale et le numéro de téléphone sur un échantillon de 1000 enregistrements WHOIS qui peut être trop petit pour obtenir de bonnes estimations permettant de comparer les régions géographiques, les bureaux d'enregistrement et les gTLD. Dans ce cadre, un échantillon plus grand devrait être envisagé dans l'avenir.

Introduction

Le 8 novembre 2012, le Conseil de l'ICANN a approuvé une série d'améliorations à la manière dont l'ICANN effectue sa surveillance du programme WHOIS (le service d'annuaire des données d'enregistrement des gTLD offert actuellement par les bureaux d'enregistrement et les opérateurs de registre), en réponse aux recommandations compilées et livrées par l'équipe de révision du WHOIS, en vertu de l'[Affirmation d'engagements](#) (AoC).

Dans le cadre de ces améliorations, l'ICANN s'est engagée à identifier de manière proactive des inexactitudes dans les données de contact des gTLD contenues dans la base WHOIS et à renvoyer ces fichiers potentiellement inexacts aux bureaux d'enregistrement afin qu'ils en assurent le suivi. Pour mener à bien ce travail, l'ICANN a entamé la mise au point du système de signalement de problèmes liés à l'exactitude du WHOIS (ARS) : un processus destiné à mettre en place des évaluations régulières de l'exactitude du WHOIS et à en publier les résultats.

L'étude pilote, conçue pour isoler et évaluer des composantes clés du système, a servi de base à l'ARS. Elle a également été conçue pour aider à assurer que l'ICANN et ses efforts soient alignés avec d'autres activités liées au WHOIS, y compris, analyser l'efficacité des nouvelles exigences en matière de validation et de vérification en vertu du contrat d'accréditation de bureau d'enregistrement (RAA), et les différents communiqués du GAC qui ont abordé l'exactitude du WHOIS.

La conception de l'ARS a été élaborée après en avoir consulté la communauté de l'ICANN au cours de la dernière année. À partir de mars 2014, l'ICANN a publié un [plan préliminaire de mise en œuvre](#) pour les ARS et pour [les commentaires publics](#), afin de demander la contribution de la communauté sur le [modèle](#) proposé qui décrit la méthodologie et l'approche. En mai 2014, suite à la clôture de la période de consultation publique, l'ICANN a publié un [appel à propositions](#) afin de chercher des fournisseurs de services pour soutenir le développement de l'ARS en utilisant une méthodologie mise à jour qui tienne compte des [commentaires](#) du public.

L'ICANN a mené l'étude pilote en août - septembre 2014 pour tester la méthodologie proposée en utilisant des données réelles et pour déterminer la faisabilité du déploiement de la technologie et des services des fournisseurs identifiés pendant le processus d'appel à propositions. [Le NORC](#)

[à l'Université de Chicago](#) a été choisi pour concevoir l'étude pilote et pour travailler avec les autres fournisseurs afin de produire un rapport concernant l'exactitude de la base des données recueillies pendant le projet pilote. Un document préliminaire des résultats a été [publié](#) avant la [réunion de l'ICANN à Los Angeles](#) pour faciliter les consultations avec la communauté de l'ICANN concernant la méthodologie, les résultats, et les prochaines étapes pour le développement de l'ARS. L'ICANN a reçu des commentaires du [GAC](#), du groupe des représentants des bureaux d'enregistrement, de l'unité constitutive des représentants de la propriété intellectuelle et d'autres sur les conclusions préliminaires du document.

Ce rapport résume les résultats de l'étude pilote, aborde les questions soulevées par la communauté lors de la réunion de l'ICANN à Los Angeles, et fournit des recommandations spécifiques pour achever la conception de l'ARS.

Conception de l'étude pilote de l'ARS

L'ARS est conçu pour examiner un enregistrement WHOIS pour chacun des trois points de vue de validation: les points de vue opérationnel, syntaxique, et de la validation de l'identité.

L'application d'une méthodologie de notation, permettra à l'ARS de publier des statistiques fondées sur certaines étiquettes d'exactitude qui seront présentées régulièrement et suivies au fil du temps. L'ARS sera limité aux enregistrements WHOIS pour les domaines génériques de premier niveau (gTLD), et exclut spécifiquement les enregistrements des domaines de premier niveau géographiques (ccTLD), comme .jp, .cn ou .uk.

Le processus proposé pour conduire l'ARS serait d'évaluer l'exactitude de l'adresse de courrier électronique, du numéro de téléphone, et de l'adresse postale d'un échantillon d'enregistrements WHOIS, selon le protocole d'évaluation établi pour chacun des points de vue d'évaluation de l'exactitude pris en compte (syntaxique, opérationnel, et l'identité possible). L'étude pilote de l'ARS explore et évalue les techniques de validation syntaxique et opérationnelle, et informera les outils et les points de vue de validation utilisés dans l'ARS. Les résultats de la validation seront ensuite analysés, et les enregistrements seront classés en fonction du degré d'exactitude.

Exigences du rapport de l'ARS

En outre, le système de signalement de problèmes liés à l'exactitude du WHOIS soutiendra l'examen d'un nombre statistiquement significatif d'enregistrements contenus dans la base WHOIS pour permettre une série de comparaisons et les rapports à publier, y compris, par exemple :

- les niveaux d'exactitude associés aux nouveaux gTLD¹ et aux gTLD plus anciens (par exemple, .com, .net, .org) ;
- Les niveaux d'exactitude associés aux enregistrements en vertu du contrat d'accréditation de bureau d'enregistrement 2013 (RAA) par rapport à la version 2009 du RAA ;

¹ Les nouveaux gTLD sont des domaines de premier niveau introduits suite à la série de candidatures menée par l'ICANN en 2012.

- Les niveaux de conformité concernant la validation et la vérification de certaines exigences relatives aux données de contact WHOIS, comme indiqué dans la spécification du programme d'exactitude du WHOIS au RAA ;
- Les niveaux d'exactitude pour les bureaux d'enregistrement, les registres et les titulaires de noms de domaine situés dans chacune des cinq régions géographiques de l'ICANN ; et
- Le classement de chaque bureau d'enregistrement accrédité par l'ICANN et le registre gTLD, en fonction des niveaux d'exactitude rapportés.

Étude pilote de l'ARS

L'étude pilote, conçue pour isoler et évaluer des composantes clés du système, servira de base à l'ARS proposé. L'étude pilote a porté sur les perspectives syntaxiques et opérationnelles de la validation des données, et en raison de la complexité et des contraintes liées au coût et au temps, n'utilise pas le point de vue de l'identité. Les principales différences entre l'étude pilote et l'ARS sont le calendrier de réalisation de l'étude, les tailles des échantillons de données WHOIS, et les points de vue pour l'évaluation de l'exactitude des données.

Pour résumer, les principaux objectifs de l'étude pilote sont les suivants :

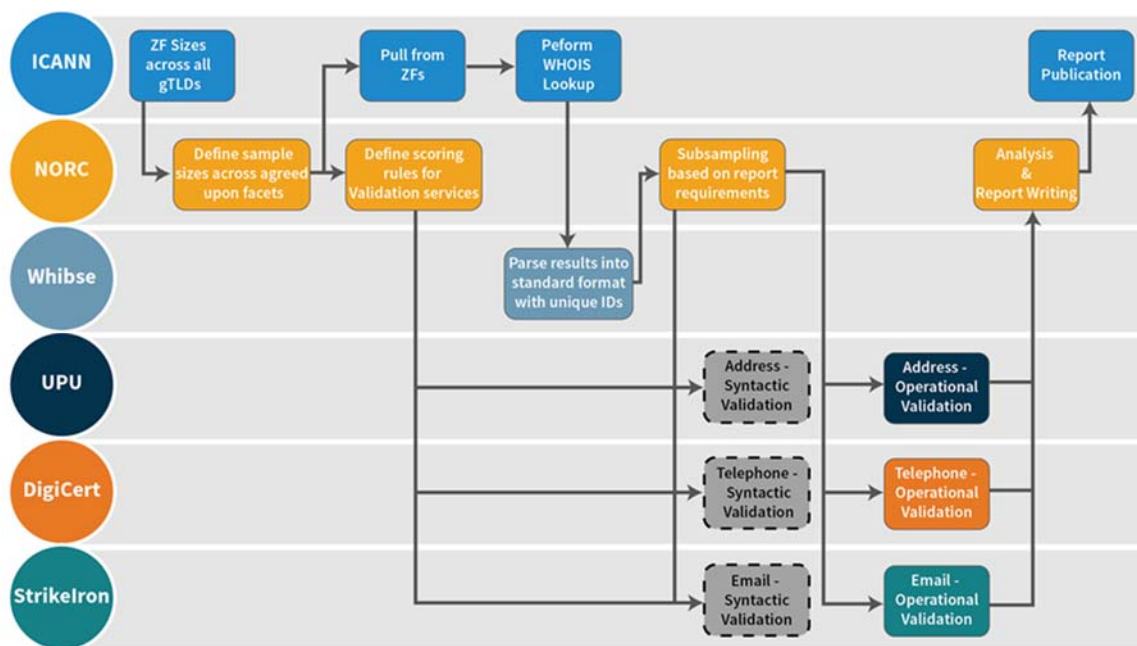
- i. informer le cadre pour la création et la réalisation de l'ARS ; et
- ii. collecter les données relatives aux taux d'exactitude des enregistrements WHOIS.

L'étude pilote a été un effort de collaboration mené par l'ICANN, avec le soutien du [NORC à l'Université de Chicago](#), [DigiCert, Inc.](#), [StrikeIron, Inc.](#), l'[Union postale universelle](#) (UPU), et [Whibse, Inc.](#). La planification initiale pour le pilote a débuté le 27 août 2014, et la collecte de données le 4 septembre 2014. La validation de l'exactitude des enregistrements WHOIS sélectionnés pour révision a été achevée pendant une période réduite, entre le 12 et 19 septembre. L'analyse des résultats de validation ont finalisé le 10 octobre 2014.

La conception de l'étude pilote a commencé par la détermination des paramètres de l'échantillon, qui a pris en compte les considérations de sous-échantillonnage et la taille de l'échantillon nécessaire pour chaque gTLD inclus dans l'étude. Un échantillon représentatif a été déterminé en utilisant les données des fichiers de zone (ZF) des gTLD. Les enregistrements sélectionnés à

partir des ZF ont ensuite subi les recherches WHOIS. Les données recueillies à partir des recherches WHOIS ont été analysées de sorte que les données puissent être interprétées de façon cohérente par les fournisseurs. Les sous-échantillons représentatifs de 10 000 et 1000 ont été sélectionnés pour la validation de l'ensemble de l'échantillon de 98 821, avec différents fournisseurs recevant différents numéros de noms de domaine à valider. Les fournisseurs ont appliqué des critères de validation prédéterminés pour l'adresse de courrier électronique, le numéro de téléphone, et les données de l'adresse postale afin d'évaluer l'exactitude du point de vue syntaxique et opérationnel. Les résultats de validation ont été compilés dans des ensembles de données, séparés par type de données et par point de vue. Ces ensembles de données ont ensuite été analysés pour produire des statistiques sur l'exactitude (pour les nouveaux gTLD et les gTLD plus anciens) stratifiées par gTLD, par région et bureau d'enregistrement. Le diagramme de flux de la conception de l'étude pilote est représenté ci-dessous dans la *Figure A*.

Figure A : diagramme de flux de la conception de l'étude pilote de l'ARS



L'approche de travail pour l'étude pilote a été divisée dans les tâches suivantes :

1. déterminer la taille de l'échantillon, le sous-échantillonnage, et la taille de l'échantillon pour chaque gTLD
2. sélectionner l'échantillon

3. analyser l'enregistrement
4. sélectionner des sous-échantillons et livrer l'analyse des ensembles de données
5. définir des règles de notation pour la validation
6. valider l'adresse de courrier électronique, le numéro de téléphone et l'adresse postale
7. analyser les ensembles de données

Tandis que les tâches 1 à 4 devaient être séquentielles, les tâches 5 à 7 pourraient être réalisées simultanément avec d'autres tâches.

Tailles des échantillons des gTLD

Le tableau A1 de l'[annexe A](#) énumère les tailles de l'échantillon pour les vingt-cinq premiers gTLD, pour les échantillons de 98 821, 10 000 et 1000. L'échantillon complet a été utilisé pour la validation syntaxique et opérationnelle de l'adresse de courrier électronique. Le sous-échantillon 10000 a été utilisé pour la validation syntaxique du numéro de téléphone et de l'adresse postale. Le sous-échantillon 1000 a été utilisé pour la validation opérationnelle du numéro de téléphone et de l'adresse postale.

Les pays du cadre d'échantillonnage

Le tableau 1 ci-dessous montre la répartition par région géographique de l'ICANN pour chacune des trois tailles d'échantillon.

Tableau 1 : tailles de l'échantillon selon la région géographique, avec le taux du sous-échantillonnage

Région géographique	Échantillon complet	Sous-échantillon 10 000	Sélection Probabilité (%)	Sous-échantillon 1 000	Probabilité de sélection (%)
Afrique (AF)	617	617	100,0	100	16,2
Amérique latine / Caraïbes (LAC)	3 380	1 000	29,6	150	15,0
Asie/Australie/Pacifique (AP)	20 176	1 784	8,9	160	8,9
Europe (EUR)	23 417	2 071	8,9	186	8,9
Amérique du nord (NA)	50 768	4 490	8,9	401	8,9
Manquant (.)	463	38	8,9	3	8,9
TOTAL / GLOBAL	98 821	10 000	10,1	1 000	10,0

Les tableaux A2a, A2b et A2c de l'[annexe A](#) ci-dessous montrent la répartition par pays. Un total de 200 pays a été représenté dans l'échantillon complet de 98 821 noms de domaine. Parmi ces 200 pays, 91 étaient représentés dans les sous-échantillons de 10 000 et de 1000 ainsi que dans l'échantillon complet (les pays d'Afrique et d'Amérique latine/Caraïbes ont été sur-échantillonnés dans les sous-échantillons). Ces 91 pays se trouvent dans le tableau A2a. Soixante-trois pays supplémentaires ont été représentés dans le sous-échantillon 10 000, mais pas au sous-échantillon 1000. Ces 63 pays se trouvent dans le tableau A2b. Les 46 pays restants sont suffisamment petits pour ne se trouver que dans l'échantillon complet, mais pas dans les sous-échantillons. Ces 46 pays se trouvent dans le tableau A2c.

Définitions d'exactitude

Pour l'étude pilote, chaque fournisseur a appliqué sa technologie et l'expertise (le cas échéant) pour analyser l'adresse postale, l'adresse de courrier électronique et le numéro de téléphone associés aux enregistrements WHOIS de l'échantillon à partir des différents points de vue. Ces enregistrements ont été divisés entre les catégories générales « exact » ou « inexact », qui regroupent les catégories de l'échelle de notation de l'exactitude de l'ARS comme suit :

Catégorie « exacte »

- Pas d'erreur : indique que les données fournies, analysées à partir des trois points de vue pour évaluer l'exactitude, contiennent les informations importantes et qu'elles sont exactes ;
- Erreur minimale : indique que les données contiennent les informations importantes. Il est possible que des clarifications s'avèrent nécessaires ou bien que des informations supplémentaires soient demandées, mais les données fournies sont exactes ; et
- Erreur limitée : indique que les données ne contiennent pas au moins une partie des informations importantes, mais que, dans une certaine mesure, les données existantes sont jugées utiles.

Catégorie « inexacte »

- Erreur majeure : indique que les données ne contiennent pas les renseignements importants, et que, pour la plupart, elles sont inexactes.
- Erreur totale : indique que les données ne contiennent pas les renseignements importants ou que les données fournies sont majoritairement inexactes.

Chaque fournisseur a ensuite déterminé comment appliquer sa technologie pour s'insérer dans cette échelle de notation. Pas toutes les sous-catégories n'ont été utilisées par les fournisseurs dans leur analyse. Pour plus d'informations sur la façon dont chaque fournisseur a utilisé cette méthode, veuillez consulter l'[annexe C](#).

Les méthodes employées par les fournisseurs au cours de l'étude pilote de l'ARS incluent des processus automatisés et manuels. Alors que l'ARS vise à automatiser au maximum le processus d'évaluation de l'exactitude, l'ICANN reconnaît que les approches automatisées ont des limitations qui doivent être complétées par l'intervention et le contrôle manuel.

Pour le fournisseur NORC à l'Université de Chicago, l'échantillonnage a impliqué des processus manuels et automatisés, tandis que l'analyse était surtout automatisée. Le processus d'échantillonnage initial a pris en compte les tailles de fichiers de zone fournis par l'ICANN. Les tailles des échantillons par zone pour l'échantillon complet ont été déterminées manuellement. Les instructions pour sélectionner des lignes dans lesquelles les fichiers ont été créés avec un outil automatisé. Les stratégies de sélection des sous-échantillons de 10 000 et 1000 ont été créées manuellement, mais le sous-échantillonnage a suivi un processus automatisé. L'analyse a été automatisée, mais quelques manuels de préparation se sont avérés nécessaires (nettoyage des données, par exemple).

L'analyse du courrier électronique menée par StrikeIron a été en grande partie automatisée, en utilisant leur technologie propriétaire pour évaluer plus de 98 000 enregistrements en quelques jours. Étant donné la facilité avec laquelle ce travail a été effectué, il pourrait être utile de continuer à utiliser un échantillon de grande taille (c'est à dire, 100 000 pour analyser les adresses électroniques dans la conception finale de l'ARS).

L'analyse du téléphone faite par DigiCert inclut divers processus manuels et automatiques. Aux cas où le code géographique du numéro de téléphone serait absent, DigiCert a déterminé manuellement si elle pourrait établir l'hypothèse du code approprié sur la base des données WHOIS restantes, avant d'examiner si le nombre approprié de numéros était présent. Il était également nécessaire de supprimer les caractères inutiles d'un numéro de téléphone, comme « + », « () », “. ” ou des espaces supplémentaires. Pour tester la validité opérationnelle du numéro de téléphone, DigiCert a utilisé un système d'appel automatisé ainsi que des personnes réelles possédant différentes compétences linguistiques pour déchiffrer les réponses reçues. La complexité de ce processus suggère l'utilisation d'un échantillon plus petit que dans le cas des courriers électroniques pour la conception finale de l'ARS.

Les services de validation de l'adresse postale menés par l'UPU consistaient également à divers processus manuels et automatiques. Étant donné que les adresses postales ont plusieurs champs, avec des normes multiples, cette analyse s'est avérée relativement complexe. En outre, il est possible de remettre des courriers électroniques à des adresses ayant certains éléments manquants. Les bases de données postales sont suffisamment sophistiquées pour déterminer si une adresse est « utilisable » dans certains endroits c'est-à-dire, si le courrier peut être livré à l'adresse, même s'il y a des champs qui manquent. Par exemple, si une adresse inclut un code postal aux États-Unis, mais la ville n'y est pas incluse, la base de données de l'UPU extrapole les informations correctes pour livrer le colis. Cependant, la qualité des bases de données postales disponibles varie d'un pays à l'autre, ce qui fait que certaines évaluations soient plus difficiles que d'autres. La complexité de ce processus suggère de sélectionner un échantillon plus petit pour les adresses postales que ce qui pourrait être réalisable pour les courriers électroniques dans la conception finale de l'ARS.

Approche et méthodes de l'étude pilote de l'ARS

Points de vue de la validation

Dans l'étude pilote, des chercheurs ont validé l'adresse de courrier électronique, le numéro de téléphone et les champs de l'adresse postale d'un sous-échantillon d'enregistrements, en examinant les données du point de vue de l'exactitude syntaxique et opérationnelle. Comme mentionné dans la Section 1 du présent rapport, l'ARS propose trois points de vue pour évaluer l'exactitude : les points de vue syntaxique, opérationnel et l'identité. Pour l'étude pilote de l'ARS, les chercheurs se sont concentrés seulement sur les points de vue syntaxique et opérationnel. Ces points de vue sont définis ci-dessous :

Point de vue syntaxique. Le point de vue syntaxique examine les données WHOIS dans le but de déterminer quelles sont les données existantes ou manquantes dans les champs de données WHOIS et évalue si ces données répondent aux normes spécifiques de formatage et de structure. La validation syntaxique offre le plus haut degré d'automatisation et donc, l'efficacité par rapport au volume dans l'évaluation de l'exactitude. Le fait de ne pas fournir une évaluation concluante de l'exactitude met en évidence la limitation de l'évaluation syntaxique qui devrait être complétée par d'autres formes de validation.

Point de vue opérationnel. Le point de vue opérationnel examine les données WHOIS dans le but de déterminer si l'information peut être pratiquement utilisée pour réaliser une fonction prévue. La validation opérationnelle combine des processus automatisés et manuels et implique la collecte et l'utilisation de données externes pour informer une évaluation de l'exactitude. Comme un compromis pour un plus faible degré d'automatisation, la validation opérationnelle fournit une notion plus tangible de l'exactitude. Bien qu'elle puisse confirmer l'applicabilité pratique des données, l'évaluation opérationnelle ne détermine pas si les données d'identification appartiennent au titulaire actuel du domaine, ce qui implique sans doute une limitation.

Point de vue de l'identité. Un troisième point de vue possible, qui n'a pas été utilisé au cours de l'étude pilote mais pour lequel l'ICANN étudie une itération ultérieure potentielle de l'ARS, examinerait les enregistrements WHOIS dans le but de confirmer les données WHOIS correspondant à la personne ou à l'entité titulaires du nom de domaine en essayant de prendre

contact avec le titulaire ou l'entité à l'aide des informations figurant dans la base de données WHOIS. L'ICANN a l'intention de travailler avec sa communauté dans les prochains mois pour étudier la faisabilité, les coûts et les avantages de l'analyse de l'exactitude du WHOIS, à partir de la perspective de l'identité.

Catégories de l'échelle de notation

Une évaluation de l'exactitude des données WHOIS, faite sous n'importe lequel des points de vue, devrait se traduire par des données étant classées selon une échelle de notation basée sur le degré auquel ces données peuvent être considérées exactes. Les catégories de l'échelle de notation initialement établies pour les études de l'exactitude du WHOIS décrivent l'exactitude comme une plage inclusive entre « pas d'erreur », le degré le plus élevé de l'exactitude et l'« erreur totale », qui représente le degré le plus bas de l'exactitude. Il existe, en tout, cinq catégories de l'échelle de notation définies comme :

- **Pas d'erreur** : indique que les données fournies contiennent les informations importantes et qu'elles sont exactes ;
- **Erreur minimale** : indique que les données contiennent les informations importantes. Il est possible que des clarifications s'avèrent nécessaires ou bien que des informations supplémentaires soient demandées, mais les données fournies sont exactes.
- **Erreur limitée** : indique que les données ne contiennent pas au moins une partie des informations importantes, mais que, dans une certaine mesure, les données existantes sont jugées utiles.
- **Erreur majeure** : indique que les données ne contiennent pas les renseignements importants, et que, pour la plupart, elles sont inexactes.
- **Erreur totale** : indique que les données ne contiennent pas les renseignements importants ou que les données fournies sont majoritairement inexactes.

Validation de l'adresse électronique

La validation syntaxique des adresses électroniques peut être décrite comme l'analyse de l'information contenue dans la syntaxe de l'adresse. Si une adresse électronique a été choisie pour

faire partie des données analysées du WHOIS, la syntaxe du courrier électronique a été vérifiée conformément aux spécifications du RFC.

La validation opérationnelle des adresses de courrier électronique a été réalisée en vérifiant le domaine dans l'adresse pour la validité et la réactivité et la tentative de communication avec les serveurs de messagerie pour déterminer si une adresse de courrier électronique était exacte et fonctionnelle. Les tentatives de communication ont retourné des informations disant que le domaine, le serveur de messagerie et l'adresse électronique pourraient être atteints, ou fournir une raison pour laquelle ils ne pourraient pas être atteints, et ces données ont été utilisées pour coder les adresses de courrier électronique dans les catégories de l'échelle de notation.

Voir l'[annexe C](#) pour une description complète du processus de validation de l'adresse de courrier électronique.

Validation du numéro de téléphone

La validation syntaxique des numéros de téléphone a été réalisée en examinant les dossiers WHOIS pour déterminer si un numéro de téléphone a été fourni et, si oui, évaluer le format du numéro fourni pour se conformer aux exigences du numéro de téléphone attendu. Le bon format a été basé sur le pays dans lequel le numéro est situé et, suite à l'évaluation, les numéros de téléphone ont été codés dans une catégorie de l'échelle de notation.

La validation opérationnelle des numéros de téléphone a été réalisée par l'évaluation de chaque numéro de téléphone afin de déterminer si un lien pourrait être établi et si des informations supplémentaires étaient nécessaires pour rendre possible une connexion. Les numéros de téléphone ont été codés en catégories de l'échelle d'évaluation en considérant si le numéro de téléphone, tel qu'il a été fourni dans le dossier WHOIS, peut se connecter avec succès, si le numéro a requis des informations supplémentaires pour se connecter ou si une connexion n'était pas possible.

Voir l'[annexe C](#) pour une description complète du processus de validation du numéro de téléphone.

Validation de l'adresse postale

La validation syntaxique des informations de l'adresse postale a été réalisée en examinant les dossiers WHOIS pour déterminer si une adresse postale a été fournie et, si oui, si le format de l'adresse était compatible avec les normes et les exigences syntaxiques attendues pour le pays dans lequel il se trouve. Lorsque les composantes d'une adresse manquent ou ne correspondent pas à des exigences syntaxiques, des contrôles supplémentaires sont réalisés pour voir si l'adresse pourrait être considérée comme utilisable. Les normes et les exigences syntaxiques attendues pour ces contrôles étaient fondées sur les directives du rapport de l'UPU de juillet 2013 sur les systèmes d'adressage postal et la conformité avec les modèles de l'UPU S42. Les adresses postales ont été codées dans une catégorie de l'échelle de notation sur la base du résultat de l'évaluation.

La validation opérationnelle des adresses postales a été réalisée en établissant si l'adresse fournie dans un dossier était une adresse existante selon la base de données des informations des adresses postales de l'UPU du monde entier. Les adresses syntaxiquement exactes, si elles étaient exactes comme prévu ou bien utilisables durant l'évaluation syntaxique, ont été incluses dans l'évaluation et chaque enregistrement a ensuite été codé dans une catégorie de l'échelle de notation. Il est important de noter que la validation des adresses postales est la plus difficile des trois validations, car plusieurs champs (numéro, rue, ville, etc.) doivent être examinés et contre-validés.

Voir l'[annexe C](#) pour une description complète du processus de validation de l'adresse postale.

Conception de l'échantillonnage

L'étude pilote a débuté avec la sélection d'un échantillon de 100 000 enregistrements des fichiers de zone des gTLD. Étant donné que les nouveaux gTLD² sont un objectif principal de l'étude, un pourcentage élevé de ces domaines ont été échantillonnés, ce qui représente 25 pour cent des domaines représentés dans l'échantillon (au moment de l'échantillonnage, seulement 1,4 pour cent des domaines appartenaient aux nouveaux gTLD). Afin de fournir des informations sur tous les gTLD, les gTLD plus petits ont été sur-échantillonnés. Plus précisément, tous les nouveaux

² Les nouveaux gTLD sont des domaines génériques de premier niveau lancés suite à la série de candidatures menée par l'ICANN en 2012.

gTLD avaient au moins 10 enregistrements sélectionnés (ou tous si le nombre de dossiers disponibles était inférieur à 10) et tous les autres gTLD (« plus anciens ») ont eu au moins 30 enregistrements sélectionnés (ou tous si le nombre de dossiers disponibles était inférieur à 30). Une fois que les enregistrements ont été sélectionnés, les enregistrements restants ont été sélectionnés proportionnellement au nombre d'enregistrements dans chaque gTLD (séparément pour les nouveaux gTLD et les gTLD plus anciens), à l'exception que l'incidence des quatre principaux nouveaux gTLD et le gTLD plus ancien ".com" ait autorisé la sélection de moins de la proportion totale de ces enregistrements pour permettre une légère augmentation de la proportion de petits gTLD. Au sein de chaque gTLD, un échantillon systématique a été spécifié (avec un enregistrement de départ et un saut d'intervalle) pour stratifier implicitement par un tri partiel ou complet dans le fichier de zone gTLD (les nouveaux domaines en haut ou en bas, par exemple). Toutes les données fournies dans ce document ne sont pas pondérées, et tandis que les petits gTLD et tous les nouveaux gTLD sont surreprésentés, les chiffres fournis ici sont à peu près représentatifs. Les poids seront calculés au fur et à mesure du progrès de l'étude dans le cadre de l'opération de l'ARS.

Une fois que les dossiers ont été sélectionnés pour l'échantillon, ils ont ensuite été interrogés par les recherches WHOIS afin de produire des données pour l'étude pilote. Une prochaine étape critique consistait à analyser les données brutes du WHOIS afin que les informations dans les dossiers puissent être systématiquement interprétées lors de l'étape de validation de l'étude. Après l'analyse, tous les enregistrements d'adresses de courrier électronique ont avancé vers les étapes de validation syntaxique et opérationnelle. Le nombre de requêtes réussies (et les enregistrements analysés) était au nombre de 98 821. Un sous-échantillon de 10 000 enregistrements a été sélectionné pour la validation syntaxique des adresses postales et des numéros de téléphone. Afin de fournir des estimations au niveau de la région géographique, NORC à l'Université de Chicago a déterminé le pays (et la région géographique) pour autant de noms de domaine que possible, et a sur-échantillonné les deux régions les plus petites : Afrique et Asie/Australie/Pacifique.

Au sein de chaque région géographique, le fichier était trié en zig-zag pour l'échantillonnage systématique afin de garantir la couverture des sous-groupes. Les variables de tri étaient :

- 1) les nouveaux gTLD contre les gTLD plus anciens ;
- 2) les bureaux d'enregistrement régis par le RAA 2009 contre ceux régis par le RAA 2013 ;
- 3) les gTLD ;
- 4) le bureau d'enregistrement ;
- 5) le pays ; et
- 6) la place dans le fichier de zone.

Le tri en zig-zag fait que les enregistrements voisins soient plus similaires que dans le cas d'un tri ordinaire (dans les nouveaux gTLD, pour exemple, les bureaux d'enregistrement régis par le RAA 2009 sont les derniers alors que pour les gTLD plus anciens, les bureaux d'enregistrement régis par le RAA 2009 sont ensemble dans la liste triée), et les résultats sont stratifiés implicitement par les variables de tri. Un petit sous-échantillon de 1000 enregistrements, (encore une fois, les régions Afrique et Asie/Australie/Pacifique ont été sur-échantillonnées) a été considéré pour la validation opérationnelle des adresses postales et des numéros de téléphone. Le même ensemble de variables a été utilisé pour classer les 10 000 enregistrements avant de sélectionner un échantillon systématique de 1000 enregistrements. Le tableau 1, ci-dessus, montre la répartition par région géographique de l'ICANN pour chacune des trois tailles d'échantillon. L'[annexe A](#) fournit les tailles d'échantillon pour les vingt-cinq gTLD les plus grands (tableau A1) et pour tous les pays avec des domaines sélectionnés (tableaux A2a, A2b, et A2c).

L'échantillon complet a été utilisé pour la validation de l'adresse de courrier électronique. La vérification syntaxique pour les numéros de téléphone et les adresses postales a utilisé le sous-échantillon de 10 000, tandis que la vérification opérationnelle pour les numéros de téléphone et les adresses postales a utilisé le sous-échantillon de 1000. Les rapports d'analyse sur les sous-groupes avec au moins 20 enregistrements, et seuls les meilleurs sept gTLD ont répondu à ce critère dans le sous-échantillon de 1000.

Il est important de noter que les différentes tailles d'échantillons se traduisent par des déterminations très différentes dans les estimations ci-dessous. Plus précisément, pour la vérification de l'adresse de courrier électronique, de petites différences sont statistiquement significatives tandis que pour la validation opérationnelle des numéros de téléphone et des

adresses postales, des différences de plusieurs points de pourcentage pourraient ne pas être statistiquement significatives. Le tableau 2 présente les erreurs type et les intervalles de confiance à 95 pour cent pour nos estimations globales de l'exactitude :

Tableau 2 : impact de la taille de l'échantillon sur les erreurs types

	Pourcentage d'exactitude : tous les gTLD	Taille de l'échantillon	Erreur standard (%)	Intervalle de confiance à 95%
Syntaxique :				
Courrier électronique	99,9	98 821	0,01	(99.7, 100.0)
Téléphone	88,7	10 000	0,32	(88.1, 89.3)
Adresse postale	81,2	10 000	0,39	(80.4, 82.0)
Opérationnel :				
Courrier électronique	89,2	98 821	0,10	(89.0, 89.4)
Téléphone	72,4	1 000	1,41	(69.6, 75.2)
Adresse postale	82,4	1 000	1,20	(80.0, 84.8)

Le tableau 2 montre que les erreurs type pour l'exactitude des adresses électroniques sont très petites, tandis que les erreurs type de l'exactitude opérationnelle pour le numéro de téléphone et l'adresse postale sont beaucoup plus grandes que les erreurs standard pour l'exactitude syntaxique. Les erreurs type pour les sous-groupes au sein de ces trois tailles d'échantillons sont plus grandes.

Le tableau 2 montre que les erreurs type pour l'exactitude des adresses électroniques sont très petites, tandis que les erreurs type de l'exactitude opérationnelle pour le numéro de téléphone et l'adresse postale sont beaucoup plus grandes que les erreurs standard pour l'exactitude syntaxique. Les erreurs type pour les sous-groupes au sein de ces trois tailles d'échantillons sont plus grandes.

Principales conclusions

Les résultats de l'analyse de l'étude pilote de l'ARS sont résumés ci-dessous. Ils montrent les statistiques du haut niveau d'exactitude des nouveaux gTLD et des gTLD plus anciens, en vertu des RAA 2009 et 2013, et par région géographique. Ces résultats sont ensuite ventilés pour montrer les taux d'exactitude des points de vue syntaxique et opérationnel pour les informations concernant l'adresse de courrier électronique, l'adresse postale et le numéro de téléphone. Dans l'ensemble, les taux d'exactitude pour les gTLD et les bureaux d'enregistrement qui avaient les échantillons les plus grands dans l'étude seront également montrés à la fin de ce résumé des résultats.

La portée de l'étude pilote n'aborde pas séparément les domaines ayant été enregistrés à travers les services d'enregistrement fiduciaire / ou d'anonymisation (PP). L'existence de ces types d'enregistrements est censée avoir un impact sur les données recueillies au cours de cette étude. L'ICANN souhaite donc s'engager avec les membres de la communauté pour discuter et identifier les éventuelles considérations PP et les impacts en prévision de futures itérations de l'ARS.

Il est vrai que, dans toute étude impliquant l'analyse de données, le cadre méthodologique, selon cette analyse, aura un grand impact sur les résultats. Par exemple, au cours de l'étude pilote de l'ARS, l'ICANN a trouvé que les évaluations de l'exactitude d'une adresse physique s'est avéré beaucoup plus difficile à réaliser que les évaluations de l'exactitude d'une adresse de courrier électronique. Et cela, en raison de la grande variété de normes de l'adresse postale et des attentes qui existent dans les différents pays. Alors que les informations de l'adresse de courrier électronique ne sont pas nécessairement plus exactes que les informations de l'adresse postale, elles sont certainement bien plus faciles à évaluer en fonction de l'effort requis. L'ICANN reconnaît que la méthodologie utilisée dans l'étude de pilote de l'ARS a été conçue pour explorer les cadres possibles pour l'évaluation de l'exactitude du WHOIS, plutôt que de fournir une solution complète pour évaluer l'exactitude.

Exactitude du point de vue de la validation

Comme indiqué dans le tableau 3 ci-dessous, l'exactitude de l'adresse électronique, le numéro de téléphone et l'adresse postale a été évaluée du point de vue syntaxique et opérationnel, et les statistiques du groupe ont été calculées pour tous les gTLD régis par les RAA 2009 et 2013 et pour les nouveaux gTLD et les gTLD plus anciens. Le tableau 3 montre les taux des enregistrements ayant été évalués comme exacts, qui incluent tous les enregistrements classés comme « pas d'erreur », « erreur minimale » et « erreur limitée ».

Tableau 3 : Pourcentages d'exactitude par point de vue, à travers les RAA, les nouveaux gTLD et les gTLD plus anciens.

	tous les gTLD	RAA 2009	RAA 2013	gTLD plus anciens	Nouveaux gTLD
Syntaxique :					
Courrier électronique	99,9	99,9	99,9	99,9	99,9
Téléphone	88,7	90,1	89,9	88,0	90,9
Adresse postale	81,2	85,9	81,0	81,8	79,1
Opérationnel :					
Courrier électronique	89,2	81,7	89,8	88,9	90,2
Téléphone	72,4	79,6	73,5	74,3	66,5
Adresse postale	82,4	85,7	82,4	82,4	82,4

Le tableau 3 montre que l'exactitude syntaxique de l'adresse de courrier électronique et du numéro de téléphone n'a pas de différence significative du point de vue statistique pour les nouveaux gTLD et les gTLD plus anciens tandis que l'exactitude syntaxique des adresses postales est meilleure pour les gTLD plus anciens. Il est à noter que même si un bureau d'enregistrement était régi par le RAA 2013 au moment de l'échantillonnage, le domaine faisant l'objet de l'échantillonnage peut avoir été enregistré lorsque le bureau d'enregistrement était régi par le RAA 2009. La validation supplémentaire et les exigences de vérification du RAA 2013 ne sont pas rétroactives pour les noms de domaine enregistrés par le bureau d'enregistrement et sont déclenchées uniquement dans les cas énumérés dans la spécification du programme de l'exactitude du Whois en vertu du RAA 2013. Les chiffres opérationnels pour les adresses de courrier électronique indiquent que l'exactitude des nouveaux gTLD est légèrement supérieure à

celle des gTLD plus anciens. L'exactitude opérationnelle des numéros de téléphone associés aux gTLD plus anciens est plus importante que celle des nouveaux gTLD mais les deux groupes affichent des résultats similaires pour l'exactitude opérationnelle des adresses postales. Une explication possible pour l'exactitude accrue des adresses de courrier électronique des nouveaux gTLD et pas des adresses postales, c'est que les enregistrements dans les nouveaux gTLD sont soumis aux exigences du RAA 2013, qui exige la vérification soit de l'adresse de courrier électronique soit du numéro de téléphone, mais pas de l'adresse postale.

Le tableau 3 montre aussi que les différences d'exactitude syntaxique entre bureaux d'enregistrement régis par le RAA 2013 et le RAA 2009 pour l'adresse de courrier électronique, numéros de téléphone et adresses postales ne sont pas statistiquement significatives. Du point de vue opérationnel, les bureaux d'enregistrement régis par le RAA 2013 affichent un taux d'exactitude des adresses de courrier électronique plus élevé que les bureaux d'enregistrement régis par le RAA 2009. Les bureaux d'enregistrement régis par le RAA 2009 ont plus d'exactitude opérationnelle pour les numéros de téléphone tandis que les deux groupes sont similaires quant aux adresses postales. Une analyse plus approfondie est nécessaire pour déterminer la raison pour laquelle les bureaux d'enregistrement régis par le RAA 2013 ont moins d'exactitude opérationnelle sur les numéros de téléphone, étant donné que les données recueillies dans l'étude pilote ne sont pas suffisantes pour explorer les raisons possibles.

Dans l'ensemble, il n'est pas surprenant que les adresses postales aient moins de précision syntaxique car plus d'un champ est impliqué. Sur le plan opérationnel, l'exactitude des numéros de téléphone est la plus faible, tandis que l'exactitude des adresses de courrier électronique est plus élevée.

Comme le tableau 3, le tableau 4 présente un rapport sur les pourcentages d'exactitude des informations de l'adresse de courrier électronique, du numéro de téléphone et de l'adresse postale par région géographique de l'ICANN.

Tableau 4 : Pourcentages d'exactitude par point de vue, dans toutes les régions géographiques de l'ICANN

	Toutes les régions géographiques	Afrique	Asie Australie Pacifique	Europe	Amérique latine et Caraïbes	Amérique du Nord
Syntactique						
Courrier électronique	99,9	99,5	99,9	99,9	99,9	99,9
Téléphone	88,7	67,4	82,6	85,9	88,9	96,1
Adresse postale	81,2	67,4	74,9	91,2	92,0	79,2
Opérationnel :						
Courrier électronique	89,2	90,9	83,2	90,0	89,8	91,1
Téléphone	72,4	52,0	57,5	65,6	76,7	85,5
Adresse postale	82,4	69,0	71,9	93,6	92,0	81,8

Le tableau 4 montre qu'en Afrique les noms de domaine ont la plus faible exactitude syntaxique dans les adresses postales et les numéros de téléphone, mais que l'exactitude syntaxique des adresses de courrier électronique est similaire dans toutes les régions géographiques. Sur le plan opérationnel, l'exactitude des domaines en Afrique et en Asie/Australie/Pacifique est plus faible pour les adresses postales et les numéros de téléphone, mais l'Afrique et l'Amérique du Nord ont la plus grande exactitude pour les adresses électroniques.

Exactitude par les gTLD

Contrairement aux tableaux ci-dessus, les tableaux 5 et 6 informent sur la validation des adresses électroniques pour les gTLD avec les dix échantillons de plus grande taille de l'étude pilote. Les noms de ces gTLD n'ont pas fait l'objet d'un rapport parce qu'il s'agit d'une étude pilote, et les pourcentages d'exactitude pour les numéros de téléphone et les adresses postales ne sont pas présentés en raison de la taille de l'échantillon. Le tableau 5 montre les pourcentages d'exactitude syntaxique des adresses de courrier électronique.

Tableau 5 : pourcentage d'exactitude syntaxique des adresses de courrier électronique pour les gTLD génériques ayant les échantillons les plus grands

TLD	Exactitude
gTLD A	100,00
gTLD B	100,00
gTLD C	100,00
gTLD D	100,00
gTLD E	99,99
gTLD F	99,95
gTLD G	99,94
gTLD H	99,92
gTLD I	99,89
gTLD J	99,88

Le tableau 6 montre les pourcentages d'exactitude opérationnelle des adresses de courrier électronique pour les gTLD dont la taille de l'échantillon est plus grande. Notez que l'ordre de ces TLD génériques est différent de celui basé sur l'exactitude syntaxique.

Tableau 6 : pourcentages de l'exactitude opérationnelle des adresses de courrier électronique pour les gTLD dont la taille de l'échantillon est plus grande

TLD	Exactitude
gTLD A	91,62
gTLD D	96,53
gTLD B	93,63
gTLD J	92,54
gTLD F	91,46
gTLD I	91,19
gTLD E	90,88
gTLD C	90,65
gTLD H	88,93
gTLD G	88,02

Exactitude par le bureau d'enregistrement

Les tableaux 7 et 8 présentent les pourcentages d'exactitude des adresses de courrier électronique par le bureau d'enregistrement. Le tableau 7 indique les pourcentages d'exactitude syntaxique des adresses de courrier électronique pour les bureaux d'enregistrement dont la taille des échantillons est plus grande.

Tableau 7 : pourcentages de l'exactitude syntaxique des adresses de courrier électronique pour les bureaux d'enregistrement dont la taille des échantillons est plus grande

Bureau d'enregistrement	Exactitude
Bureau d'enregistrement A	100,00
Bureau d'enregistrement B	100,00
Bureau d'enregistrement C	100,00
Bureau d'enregistrement D	100,00
Bureau d'enregistrement E	100,00
Bureau d'enregistrement F	99,97
Bureau d'enregistrement G	99,95
Bureau d'enregistrement H	99,95
Bureau d'enregistrement I	99,93
Bureau d'enregistrement J	99,92

Le tableau 8 montre les pourcentages d'exactitude opérationnelle pour les adresses de courrier électronique pour les bureaux d'enregistrement dont la taille des échantillons est plus grande. Comme dans le cas des gTLD avec la plus grande taille de l'échantillon, l'ordre de l'exactitude opérationnelle pour les bureaux d'enregistrement change par rapport à l'ordre de l'exactitude syntaxique.

Tableau 8 : pourcentage d'exactitude opérationnelle des adresses de courrier électronique pour les bureaux d'enregistrement dont la taille des échantillons est plus grande

Bureau d'enregistrement	Exactitude
Bureau d'enregistrement F	94,6
Bureau d'enregistrement D	94,4
Bureau d'enregistrement I	94,0
Bureau d'enregistrement A	94,0
Bureau d'enregistrement E	93,9
Bureau d'enregistrement C	92,6
Bureau d'enregistrement G	92,6
Bureau d'enregistrement H	92,5
Bureau d'enregistrement B	92,3
Bureau d'enregistrement J	89,6

Conséquences des résultats

L'étude pilote a été un exercice très instructif, et a fourni de précieuses leçons et mises en garde pour les méthodes et procédures de l'ARS. Quelques-unes parmi les principales leçons tirées du projet pilote sont détaillées ci-dessous.

Examen des définitions

Dans les mois à venir, l'ICANN affinera les éléments de l'étude qui seront critiques pour l'établissement d'un cadre pouvant être reproduit pour la mise en œuvre de l'ARS. Les modalités de détermination des enregistrements WHOIS exacts et inexacts doivent être examinées davantage. Les fournisseurs qui ont effectué les étapes de validation ont fait des évaluations raisonnables de l'exactitude de l'adresse postale, de l'adresse électronique et du numéro de téléphone. Toutefois, il faut réfléchir davantage pour savoir si les évaluations s'adaptent aux catégories d'exactitude (pas d'erreur, erreur minimale, erreur limitée, erreur majeure, erreur totale). Une évaluation des méthodologies sera effectuée afin de déterminer les moyens pour que les évaluations de l'exactitude du téléphone, de l'adresse électronique et de l'adresse postale soient comparables.

Un aspect clé de la conception de l'ARS comprend un examen minutieux de la méthodologie qui sera déployée pour chaque point de vue, afin d'appliquer l'échelle de notation. Cela suppose d'interroger les codes d'erreur qui peuvent être générés par la technologie du fournisseur pour confirmer comment il affecte l'exactitude. Par exemple, un code d'erreur marquant qu'une adresse de courrier électronique est associée à un site Web malveillant devrait-il affecter l'exactitude de la base de données WHOIS ? Si un serveur de messagerie est configuré pour « attraper tous » les courriers électroniques envoyés au domaine, indépendamment de savoir si l'adresse électronique spécifique a été créée, cela doit-il être considéré exact ou inexact du point de vue opérationnel ?

L'ICANN est intéressée à recevoir la collaboration de la communauté sur ces aspects de la conception de l'ARS au cours de la période de consultation publique sur le rapport pilote.

Déterminer un pointage pour l'exactitude

Bien que l'on puisse beaucoup apprendre de l'analyse des pourcentages d'exactitude des adresses de courrier électronique, des numéros de téléphone et des adresses postales, une notation du point de vue de l'exactitude et une notation globale à travers les différents points de vue fournirait davantage d'informations sur l'état de l'exactitude des enregistrements WHOIS. Ci-dessous sont décrites les grandes lignes d'une méthodologie de notation qui utiliserait l'évaluation de l'exactitude de chaque enregistrement échantillonné pour déterminer un point de vue (syntaxique, opérationnel) et la notation globale de l'exactitude. Ces notations des enregistrements peuvent être utilisées pour analyser l'état de l'exactitude des groupes d'enregistrements (par gTLD, pays, etc.).

Comme dans l'étude pilote, les champs de l'adresse de courrier électronique, du numéro de téléphone et de l'adresse postale de chaque enregistrement WHOIS échantillonné sont évalués afin de déterminer si le champ est exact ou inexact. Rappelons que le champ de l'enregistrement est considéré comme exact s'il est noté comme erreur limitée, erreur minimale ou pas d'erreur. Le champ de l'enregistrement est considéré comme inexact s'il est noté comme erreur majeure ou erreur totale. La détermination exact/inexact peut être convertie en une valeur numérique en attribuant la valeur 1 à un champ exact et -1 à un champ inexact. La notation du point de vue syntaxique ou opérationnel de l'enregistrement est la somme de la valeur de l'exactitude de l'adresse de courrier électronique, du numéro de téléphone et de l'exactitude. Comme indiqué dans le tableau 9, la notation de l'exactitude de l'enregistrement sera de -3, -1, 0, 1 ou 3.

Tableau 9 : notation du point de vue de l'enregistrement

Valeur de l'exactitude			Notation du point de vue de l'exactitude des enregistrements
Adresse électronique	Numéro de téléphone	Adresse postale	
-1	-1	-1	-3
-1	-1	1	-1
-1	1	-1	
1	-1	-1	
1	1	-1	1
1	-1	1	
-1	1	1	
1	1	1	3

Les notations du point de vue de l'exactitude peuvent être agrégées dans un groupe d'enregistrements. Le pourcentage d'enregistrements dans le groupe ayant des notations de -3, -1, 1 ou 3 peut être calculé, et ces pourcentages peuvent être comparés entre les groupes. Par exemple, on pourrait calculer le pourcentage d'enregistrements avec des notes d'exactitude de 1 ou 3 pour les nouveaux gTLD et le comparer à un pourcentage similaire pour les gTLD plus anciens. En outre, on pourrait faire la moyenne entre la notation de tous les enregistrements appartenant à un groupe afin d'obtenir une notation de groupe dont les valeurs seront comprises entre -3 et 3. Les notations de groupe peuvent également être comparées statistiquement entre les différents groupes d'intérêt. En outre, la notation de groupe peut être associée aux cinq catégories de l'échelle de notation de l'exactitude. Le tableau 10 fournit une cartographie proposée.

Tableau 10 : échelle de notation de l'exactitude proposée pour la cartographie de la notation de groupe

Échelle de notation de l'exactitude	Plage de notation de groupe
Pas d'erreur	2 – 3
Erreur minimale	0,5 – 2
Erreur limitée	-0,5 – 0,5
Erreur majeure	-2 – -0,5
Erreur totale	-3 – -2

La note globale de l'exactitude de l'enregistrement peut être déterminée en faisant la moyenne des notations du point de vue de l'enregistrement. Ensuite, on peut faire la moyenne de la notation globale de l'enregistrement entre un groupe d'enregistrements pour donner une notation moyenne au groupe d'intérêt. Des comparaisons statistiques entre les groupes peuvent être considérées comme pour les notations en perspective.

Taille de l'échantillon et niveaux de confiance

La validation des adresses postales et des numéros de téléphone prend plus de temps que celle des adresses de courrier électronique. En outre, la validation opérationnelle prend plus de temps que la validation syntaxique des adresses postales et des numéros de téléphone. Voilà la raison pour laquelle des échantillons de différentes tailles ont été utilisés. Cela signifie que les mesures de l'exactitude effectuées sur la base d'échantillons plus petits sont moins fiables que celles effectuées sur des échantillons plus grands. En particulier, la validation opérationnelle pour l'adresse postale et le numéro de téléphone sur un échantillon de seulement 1000 enregistrements WHOIS Le tableau 2 montre les erreurs standard pour les 1000 enregistrements de plus de 1 pour cent. Pour les sous-groupes plus petits tels que les régions géographiques, les bureaux d'enregistrement et les gTLD, les erreurs standard sont beaucoup plus nombreuses, et en conséquence, 1000 est trop peu pour fournir de bonnes estimations permettant de comparer les régions géographiques, les bureaux d'enregistrement et les gTLD. Dans ce cadre, un échantillon plus grand devrait être envisagé dans l'avenir. Nous suggérons qu'une taille minimale optimale pour toute analyse devrait être de 5000 ou 10 000, mais l'augmentation de la taille minimale de 1000 à 2000 augmenterait considérablement la valeur de l'analyse des données.

Pas de rapport sur la conformité

Le rapport pilote n'était pas censé être un rapport visant à évaluer la conformité avec les exigences de validation et vérification du nouveau RAA 2013. Par contre, il a examiné l'exactitude à partir de différents points de vue, comme décrit ci-dessus. Dans la conception finale de l'ARS, l'ICANN prévoit de développer et de publier des rapports complémentaires visant à respecter plus étroitement les exigences du RAA 2013.

Explorer la validation de l'identité

L'étude pilote ne comportait pas une évaluation de l'exactitude du point de vue de l'identité. Dans les prochains mois, l'ICANN envisage d'étudier la possibilité de concevoir l'ARS pour inclure l'examen d'un enregistrement WHOIS du point de vue de l'identité.

PROCHAINES ETAPES

L'ICANN prévoit de publier ce rapport pour commentaires du public pendant une période qui sera clôturée le 27 février 2015. Au cours de la période de commentaires publics, l'ICANN demande des commentaires sur :

1. le plan d'étude et la méthodologie
2. les types de rapports d'exactitude qui seront publiés à travers l'ARS
3. si l'ICANN devrait effectuer des vérifications d'identité dans les étapes ultérieures du développement de l'ARS
4. déterminer si la méthodologie devrait traiter différemment les enregistrements sous les services d'anonymisation ou d'enregistrement fiduciaire et dans l'affirmative, comment
5. tout autre aspect de l'ARS

Cette rétroaction sur le rapport du projet pilote au cours de cette période de consultation publique, ainsi que par le dialogue de la communauté lors de la [52 réunion de l'ICANN à Singapour](#) en février 2015, informera le personnel de l'ICANN et contribuera à façonner le développement futur du système de signalement de problèmes liés à l'exactitude du WHOIS.

L'[annexe B](#) contient des informations supplémentaires de l'ICANN concernant les prochaines étapes pour finaliser le développement de l'ARS.

ANNEXE A. Taille des échantillons par gTLD et par pays.

Tableau A1 : taille des échantillons pour les vingt-neuf gTLD principaux

gTLD	Échantillon complet	Sous-échantillon 10 000	Sélection Probabilité (%)	Sous-échantillon 1 000	Probabilité de sélection (%)
.com	44 399	4 518	10,2	451	10,0
.net	12 539	1 261	10,1	129	10,2
.org	8 726	883	10,1	89	10,1
.info	4 725	456	9,7	44	9,6
.XYZ (nouveau)	3 282	293	8,9	26	8,9
.biz	2 215	223	10,1	24	10,8
.berlin (nouveau)	1 504	134	8,9	12	9,0
.club (nouveau)	1 162	141	12,1	17	12,1
.guru (nouveau)	805	82	10,2	10	12,2
.mobi	781	84	10,8	9	10,7
.wang (nouveau)	729	64	8,8	6	9,4
.photography (nouveau)	514	52	10,1	6	11,5
.email (nouveau)	457	49	10,7	7	14,3
.link (nouveau)	439	48	10,9	5	10,4
.xn--3ds443g (nouveau)	420	38	9,0	3	7,9
.xn--55qx5d (nouveau)	420	37	8,8	4	10,8
.today (nouveau)	419	47	11,2	5	10,6
.tips (nouveau)	343	37	10,8	5	13,5
.company (nouveau)	330	36	10,9	3	8,3
.xn--io0a7i (nouveau)	296	27	9,1	2	7,4
.solutions (nouveau)	291	29	10,0	3	10,3
.center (nouveau)	265	28	10,6	4	14,3
.tokyo (nouveau)	255	23	9,0	2	8,7
.asia	253	25	9,9	1	4,0
.expert (nouveau)	247	25	10,1	3	12,0

Tableau A2c : taille des échantillons pour les 91 pays dans le sous-échantillon 1000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon	Sous-échantillon 10 000	Sous-échantillon 1 000
Manquant	463	0,47	38	3
États-Unis (US)	46 995	47,65	4151	367
Chine (CN)	7 347	7,45	653	57
Allemagne (DE)	6 402	6,49	570	43
Royaume-Uni (UK)	4 779	4,85	431	42
Canada (CA)	3 711	3,76	334	34
Japon (JP)	3 349	3,40	298	30
Australie (AU)	3 103	3,15	277	22
France (FR)	2 296	2,33	192	16
Panama (PA)	1 661	1,68	491	77
Espagne (ES)	1 511	1,53	147	19
Pays-Bas (NL)	1 237	1,25	103	8
Turquie (TR)	1 204	1,22	104	8
Inde (IN)	1 123	1,14	100	5
Italie (IT)	939	0,95	78	7
Îles Caïman (EU)	880	0,89	80	9
Suisse(CH)	792	0,80	70	4
Nouvelle-Zélande (NZ)	780	0,79	67	7
Fédération de Russie (RU)	569	0,58	60	3
Autriche (AT)	497	0,50	37	2
Brésil (BR)	485	0,49	149	26
Corée, République de (KR)	484	0,49	47	4
Suède (SE)	353	0,36	32	5
Mexique (MX)	338	0,34	99	16

Tableau A2c : taille des échantillons pour les 91 pays dans le sous-échantillon 1000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon	Sous-échantillon 10 000	Sous-échantillon 1 000
Hong Kong (HK)	310	0,31	25	6
Belgique (BE)	295	0,30	24	3
Indonésie (ID)	290	0,29	27	3
Vietnam (VN)	288	0,29	23	1
Irlande (IE)	283	0,29	26	4
Pologne (PL)	281	0,28	27	3
Thaïlande (TH)	279	0,28	22	3
Norvège (NO)	264	0,27	24	2
Afrique du Sud (ZA)	223	0,23	223	35
Danemark (DK)	196	0,20	19	1
Finlande (FI)	178	0,18	16	2
Singapour (SG)	174	0,18	18	3
Bermudes (BM)	170	0,17	14	1
Émirats arabes unis (AP)	166	0,17	15	2
Malaisie (MY)	165	0,17	16	1
Ukraine (UA)	156	0,16	14	1
Portugal (PT)	136	0,14	12	2
Argentine (AR)	132	0,13	42	4
Colombie (CO)	126	0,13	39	4
Bulgarie (BG)	112	0,11	11	3
Grèce (GR)	111	0,11	7	1
Arabie Saoudite (SA)	110	0,11	10	1
Taiwan (TW)	100	0,10	8	1
Philippines (PH)	97	0,10	6	1
Égypte (EG)	91	0,09	91	16

Tableau A2c : taille des échantillons pour les 91 pays dans le sous-échantillon 1000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon	Sous-échantillon 10 000	Sous-échantillon 1 000
Pakistan (PK)	87	0,09	6	1
Pérou (PE)	80	0,08	24	3
Antigua-et-Barbuda (AG)	78	0,08	22	2
Nigéria (NG)	73	0,07	73	10
Chypre (CY)	59	0,06	5	1
Luxembourg (LU)	58	0,06	5	1
Costa Rica (CR)	56	0,06	15	2
Venezuela (VE)	55	0,06	12	1
Chili (CL)	52	0,05	16	1
Îles Vierges britanniques (VG)	48	0,05	5	1
Saint-Christophe-et-Niévès (KN)	46	0,05	13	2
Liban (LB)	45	0,05	7	1
Maroc (MA)	38	0,04	38	5
Équateur (EC)	36	0,04	10	3
Sri Lanka (LK)	25	0,03	1	1
République Dominicaine (DO)	24	0,02	5	2
Biélorussie (BY)	23	0,02	3	1
Kenya (KE)	22	0,02	22	4
Uruguay (UY)	22	0,02	8	2
Guatemala (GT)	21	0,02	8	1
Lituanie (LT)	17	0,02	2	1
Seychelles (SC)	16	0,02	16	2
Lettonie (LV)	16	0,02	1	1

Tableau A2c : taille des échantillons pour les 91 pays dans le sous-échantillon 1000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon	Sous-échantillon 10 000	Sous-échantillon 1 000
Ghana (GH)	15	0,02	15	4
Namibie (NA)	14	0,01	14	3
Bahamas (BS)	14	0,01	5	3
Ouganda (UG)	13	0,01	13	3
Algérie (DZ)	13	0,01	13	1
St Vincent et Grenad. (VC)	13	0,01	3	1
Tunisie (TN)	11	0,01	11	4
Swaziland (SZ)	10	0,01	10	1
Ouzbekistan (UZ)	7	0,01	1	1
Rép. démocratique du Congo (CD)	6	0,01	6	1
Cote d'Ivoire (CI)	6	0,01	6	1
Sahara occidental (EH)	6	0,01	6	1
Zimbabwe (ZW)	5	0,01	5	1
Botswana (BW)	4	0,00	4	2
Burkina Faso (BF)	2	0,00	2	1
Bénin (BJ)	2	0,00	2	1
Madagascar (MG)	2	0,00	2	1
Mali (ML)	2	0,00	2	1
Angola (AO)	1	0,00	1	1
Cap-Vert (CV)	1	0,00	1	1

Tableau A2b : taille des échantillons pour tous les autres 63 pays dans le sous-échantillon 10 000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon (%)	Sous-échantillon 10 000
Israël (IL)	173	0,18	12
République tchèque (CZ)	165	0,17	14
Roumanie (RO)	109	0,11	10
Curaçao (CW)	83	0,08	6
République islamique de l'Iran (IR)	79	0,08	7
Hongrie (HU)	71	0,07	6
Yougoslavie (Serbie et Monténégro) (YU)	50	0,05	2
Slovaquie (SK)	49	0,05	8
Croatie (HR)	46	0,05	4
Afghanistan (AF)	39	0,04	2
Bangladesh (BD)	37	0,04	2
Jordanie (JO)	36	0,04	3
Koweït (KW)	34	0,03	5
Puerto Rico (PR)	31	0,03	3
Barbados (BB)	30	0,03	12
Malte (MT)	30	0,03	1
Slovénie (SI)	30	0,03	1
Trinité-et-Tobago (TT)	21	0,02	6
Azerbaïdjan (AZ)	21	0,02	2
Estonie (EE)	20	0,02	1
Belize (BZ)	18	0,02	5
Bolivie (BO)	18	0,02	3

Tableau A2b : taille des échantillons pour tous les autres 63 pays dans le sous-échantillon 10 000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon (%)	Sous-échantillon 10 000
Gibraltar (GI)	18	0,02	1
Cambodge (KH)	18	0,02	1
Jamaïque (JM)	16	0,02	5
Macédoine, ex-République yougoslave (MK)	13	0,01	1
Qatar (QA)	12	0,01	2
El Salvador (SV)	11	0,01	3
Guadeloupe (GP)	11	0,01	1
Népal (NP)	11	0,01	1
Islande (IS)	10	0,01	1
Tanzanie, Rép. Unie. TZ	9	0,01	9
Paraguay (PY)	8	0,01	2
Géorgie (GE)	8	0,01	1
Palaos (PW)	8	0,01	1
Iraq (IQ)	7	0,01	2
Birmanie (MM)	7	0,01	2
Macao (MO)	7	0,01	2
Samoa américaine (AS)	7	0,01	1
Brunei Darussalam (BN)	7	0,01	1
Îles Mariannes du Nord (MP)	7	0,01	1
Île Maurice (MU)	6	0,01	6
Andorre (AD)	6	0,01	1
Arménie (AM)	6	0,01	1
Saint-Pierre et Miquelon (PM)	6	0,01	1

Tableau A2b : taille des échantillons pour tous les autres 63 pays dans le sous-échantillon 10 000

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon (%)	Sous-échantillon 10 000
Sénégal (SN)	5	0,01	5
Nicaragua (NI)	5	0,01	2
Niger (NE)	4	0,00	4
Guernsey (GG)	4	0,00	1
Éthiopie (ET)	3	0,00	3
Îles Turques-et-Caïques (TC)	3	0,00	1
Cameroun (CM)	2	0,00	2
Soudan (SD)	2	0,00	2
Somalie (SO)	2	0,00	2
Zambie (ZM)	2	0,00	2
Dominique (DM)	2	0,00	1
Burundi (BI)	1	0,00	1
Bhoutan (BT)	1	0,00	1
Congo (CG)	1	0,00	1
Djibouti (DJ)	1	0,00	1
Jamahiriya arabe libyenne (LY)	1	0,00	1
Malawi (MW)	1	0,00	1
Togo (TG)	1	0,00	1

Tableau A2c : taille des échantillons pour tous les autres 46 pays dans le sous-échantillon complet

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon (%)	Sous-échantillon 10 000
Albanie (AL)	16	0,02	0
Liechtenstein (LI)	12	0,01	0
Bosnie-Herzégovine (BA)	11	0,01	0
Barheïn (BH)	10	0,01	0
Kazakhstan (KZ)	10	0,01	0
Monaco (MC)	10	0,01	0
Île Norfolk (NF)	10	0,01	0
Île de Man (IM)	9	0,01	0
Îles Mineures des États-Unis (UM)	8	0,01	0
Îles Cocos (Keeling) (CC)	7	0,01	0
Moldavie, République de (MD)	7	0,01	0
San Marino (SM)	7	0,01	0
Terres australes françaises (TF)	6	0,01	0
Îles Vierges américaines (VI)	6	0,01	0
Honduras (HN)	5	0,01	0
Réunion (RE)	5	0,01	0
Vanuatu (VU)	5	0,01	0
Maldives (MV)	4	0,00	0
Oman (OM)	4	0,00	0
Guam (GU)	3	0,00	0
Mongolie (MN)	3	0,00	0
Polynésie française (PF)	3	0,00	0
Papouasie-Nouvelle-Guinée (PG)	3	0,00	0

Tableau A2c : taille des échantillons pour tous les autres 46 pays dans le sous-échantillon complet

Pays	Échantillon complet (98 821)	Pourcentage de l'échantillon (%)	Sous-échantillon 10 000
Territoires palestiniens (PS)	3	0,00	0
Yémen (YE)	3	0,00	0
Jersey (JE)	2	0,00	0
Kyrgyzstan (KG)	2	0,00	0
Sainte Lucie (LC)	2	0,00	0
Îles Marshall (MH)	2	0,00	0
Suriname (SR)	2	0,00	0
Île de l'Ascension (AC)	1	0,00	0
Aruba (AW)	1	0,00	0
Cuba (CU)	1	0,00	0
Île Christmas (CX)	1	0,00	0
Grenade (GD)	1	0,00	0
Groenland (GL)	1	0,00	0
Haïti (HT)	1	0,00	0
République démocratique de Corée (KP)	1	0,00	0
République populaire de LAO (LA)	1	0,00	0
Nouvelle Calédonie (AP)	1	0,00	0
Sainte-Hélène (EU)	1	0,00	0
République arabe syrienne (SY)	1	0,00	0
Tadjikistan (TJ)	1	0,00	0
Timor-Leste (TL)	1	0,00	0
Tuvalu (TV)	1	0,00	0
Wallis-et-Futuna (WF)	1	0,00	0

Annexe B : Prochaines étapes pour l'élaboration du système de signalement de problèmes liés à l'exactitude du WHOIS (ARS)

Ce document est publié par l'ICANN comme accompagnement du rapport pilote final d'exactitude de NORC (rapport NORC) pour décrire les prochaines étapes pour le développement du système de signalement de problèmes liés à l'exactitude du WHOIS (ARS).

Commentaires et consultation publics à Singapour

Tel que décrit dans le rapport NORC, un forum de commentaires publics sera ouvert jusqu'au 27 février 2015. En outre, l'ICANN envisage d'entrer en contact avec les parties prenantes concernées lors de la [52e réunion de l'ICANN à Singapour](#) en février 2015 afin d'obtenir des retours sur la conception proposée pour l'ARS tel que décrite dans le rapport NORC. Après avoir examiné ces observations et commentaires, l'ICANN mettra à jour la conception de l'ARS et commencera son développement par étapes, comme décrit ci-dessous.

Au cours de la consultation publique, l'ICANN vise à obtenir des commentaires sur :

1. le plan d'étude et la méthodologie
2. les types de rapports d'exactitude qui seront publiés à travers l'ARS
3. si l'ICANN devait effectuer des vérifications d'identité dans les étapes ultérieures du développement de l'ARS
4. déterminer si la méthodologie devrait traiter différemment les enregistrements sous les services d'anonymisation ou d'enregistrement fiduciaire et dans l'affirmative, comment
5. tout autre aspect de l'ARS

a. Étape 1 – exactitude syntaxique

L'étape 1 se concentrera sur le rapport des niveaux d'exactitude qui examinent les aspects syntaxiques du courrier électronique, du numéro de téléphone et des adresses postales. Vu que ces examens peuvent être effectués en utilisant des processus largement automatisés, il est prévu

que cette partie de l'ARS puisse être lancée au début ou vers la mi-2015. Un **plan de mise en œuvre final - étape 1** sera élaboré pour spécifier la taille, le processus et la méthodologie de classification des échantillons qui seront mis en œuvre pour examiner l'exactitude du WHOIS du point de vue syntaxique.

b. **Étape 2 – exactitude opérationnelle**

Ensuite, l'ICANN prévoit de développer davantage l'ARS pour signaler des niveaux d'exactitude qui examinent les enregistrements WHOIS du point de vue opérationnel, en ce qui concerne le courrier électronique, le numéro de téléphone et les adresses postales. Étant donné que cet examen implique de longs processus manuels, il est probable que des échantillons de plus petite taille soient déployés. L'ICANN élaborera un **plan de mise en œuvre final - étape 2**, pour tenir compte des leçons apprises au cours de l'étape 1, pour un lancement vers la mi-fin 2015.

c. **Étape 3 – concept d'exactitude du point de vue de l'identité**

L'étape finale examine la possibilité et la manière d'effectuer des études d'exactitude permanentes pour ce qui est de la confirmation de l'identité du titulaire. Le Comité consultatif gouvernemental (GAC) dans son [communiqué de Los Angeles](#) a conseillé à l'ICANN de prendre des mesures pour encadrer et examiner les risques, la faisabilité, les coûts et les avantages de la réalisation d'études permanentes d'exactitude pour valider et vérifier l'identité du titulaire. Le personnel est en train d'élaborer une réponse au GAC qui fournira cette information supplémentaire en janvier, avant la [52e réunion de l'ICANN à Singapour](#). L'étape 3 de l'ARS évaluera la faisabilité et les coûts de la réalisation des études d'exactitude basées sur l'identité du titulaire.

Processus de correction des enregistrements inexacts : groupe consultatif pour la mise en œuvre de l'ARS

Une des fonctions essentielles de l'ARS sera de transmettre les enregistrements identifiés comme potentiellement inexacts aux bureaux d'enregistrement pour qu'ils fassent un suivi et confirment leur exactitude. L'ARS est conçu pour faire le suivi du progrès de ces enregistrements.

En janvier 2015, l'ICANN lancera une étude pilote de la conformité afin d'examiner les résultats du projet pilote et déterminer si une réponse de conformité est appropriée pour les enregistrements WHOIS qui ont été catégorisés comme inexacts du point de vue syntaxique. Le département de la conformité contractuelle de l'ICANN fait constamment des audits des résultats de l'étude pilote dans le cadre d'un pilote de conformité afin de déterminer si un suivi de conformité est nécessaire.

En outre, l'ICANN compte entrer en rapport avec les bureaux d'enregistrement et d'autres parties intéressées au besoin pour définir la façon d'intégrer la transmission, la révision et la mise à jour, le cas échéant, des enregistrements WHOIS qui ont été identifiés comme potentiellement inexacts dans les systèmes et les processus de conformité existants. L'ICANN prévoit de travailler avec les bureaux d'enregistrement et la communauté élargie dans les mois à venir afin de développer et de peaufiner ce processus.

Annexe C : DESCRIPTIONS DÉTAILLÉES DES PROCESSUS DES FOURNISSEURS

Pour l'étude pilote, chaque fournisseur a appliqué sa technologie et son expertise (le cas échéant) pour analyser l'adresse postale, l'adresse de courrier électronique et le numéro de téléphone associés aux enregistrements WHOIS de l'échantillon des différents points de vue. Ces enregistrements ont été divisés entre les catégories générales « exact » ou « inexact », qui regroupent les catégories de l'échelle de notation de l'exactitude de l'ARS comme suit :

Catégorie « exacte »

- Pas d'erreur : indique que les données fournies, analysées à partir des trois points de vue pour évaluer l'exactitude, contiennent les informations importantes et qu'elles sont exactes ;
- Erreur minimale : indique que les données contiennent les informations importantes. Il est possible que des clarifications s'avèrent nécessaires ou bien que des informations supplémentaires soient demandées, mais les données fournies sont exactes ; et
- Erreur limitée : indique que les données ne contiennent pas au moins une partie des informations importantes, mais que, dans une certaine mesure, les données existantes sont jugées utiles.

Catégorie « inexacte »

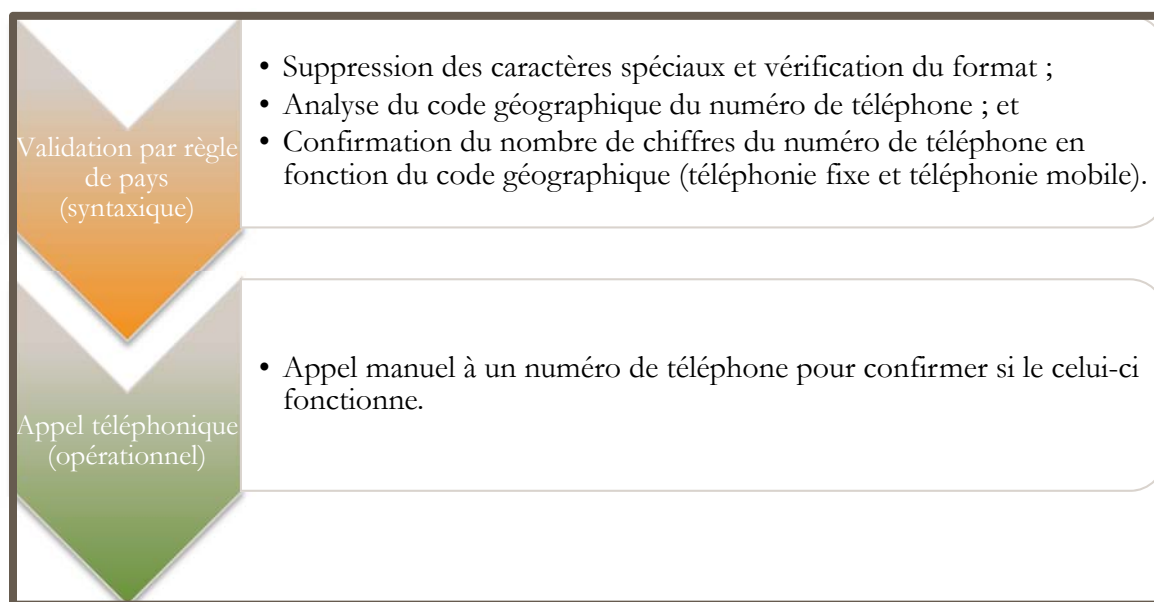
- Erreur majeure : indique que les données ne contiennent pas les renseignements importants, et que, pour la plupart, elles sont inexactes.
- Erreur totale : indique que les données ne contiennent pas les renseignements importants ou que les données fournies sont majoritairement inexactes.

Cette annexe fournit des explications détaillées des processus suivis par chaque fournisseur pour les éléments de données spécifiques analysés ainsi qu'une vue d'ensemble de la manière dont chaque fournisseur a appliqué son analyse à l'échelle de notation NORC.

Validation du téléphone : DigiCert

L'ICANN a passé un contrat avec DigiCert pour assurer l'évaluation de l'exactitude WHOIS du téléphone.

Le tableau suivant décrit brièvement les mesures générales prises par DigiCert, avec une explication détaillée ci-dessous.



Validation syntaxique du téléphone

Une fois que les enregistrements analysés ont été reçus par DigiCert, l'étape initiale a été de « nettoyer » les données de téléphone de l'enregistrement pour supprimer les caractères spéciaux (parenthèses, plus, points, etc.) et leur donner le format correct.

Une fois que cet examen initial a été achevé, DigiCert a vérifié chaque enregistrement pour déterminer si un code géographique pouvait être identifié dans le numéro de téléphone. Dans l'affirmative, DigiCert a ensuite analysé le code géographique des enregistrements et commencé l'étape suivante de la vérification.

Si l'enregistrement avait un code géographique dans le champ du titulaire, DigiCert a comparé les chiffres restants dans le numéro de téléphone pour voir s'il contenait la bonne quantité de chiffres, que ce soit un numéro de téléphonie fixe ou de téléphonie mobile, pour le pays. Pour déterminer le nombre approprié de chiffres pour un pays, DigiCert a invoqué la recommandation E.164 du secteur de normalisation des télécommunications de l'UIT (UIT-T E.164).

En outre, DigiCert n'a pas tenté d'effectuer une validation croisée des numéros de téléphone et des adresses postales. Par exemple, si un numéro de téléphone contenait un code géographique et la quantité de chiffres correspondante à ce code géographique, le code géographique n'a pas fait l'objet d'une validation croisée avec l'adresse. Puisqu'il n'y a aucune exigence de faire

correspondre les pays des numéros de téléphone avec les adresses postales d'un enregistrement WHOIS, les adresses postales ont été référencées seulement s'il était nécessaire de « supposer » un code géographique pour un enregistrement téléphonique qui manquait les informations de préfixe du code géographique.

Un dossier a reçu une notation de « pas d'erreur (+ 2) » si le numéro de téléphone était présent dans les données WHOIS et il contenait un nombre acceptable de chiffres sous les exigences des numéros de téléphone applicables au code géographique pertinent. L'enregistrement devait inclure un préfixe de code géographique pour être noté (+ 2).

Un enregistrement a reçu une notation « erreur minimale (+ 2) » s'il y avait des erreurs de mise en forme dans l'enregistrement téléphonique. Toutefois, le numéro de téléphone contenait le nombre acceptable de chiffres pour le code géographique. Par conséquent, DigiCert a ignoré les problèmes de mise en forme et a essentiellement appliqué une notation « pas d'erreur (+ 2) » à l'enregistrement. Un exemple en serait l'inclusion de (“.”) ou (“+”) dans les données.

Un enregistrement a reçu une notation « erreur limitée (+ 0) » si le numéro de téléphone manquait le code géographique mais contenait un nombre acceptable de chiffres, si le code géographique correspondant au pays identifié dans le champ d'adresse du contact spécifique (c'est à dire, le contact administratif) a été ajouté à l'enregistrement.

Un enregistrement a reçu une notation « erreur majeure (-1) » si le numéro de téléphone du contact administratif ou technique ne contenait pas le code géographique mais le champ d'adresse de contact du titulaire contenait un pays. DigiCert a ensuite déterminé si le numéro de téléphone, avec le code géographique du titulaire ajouté grâce à l'adresse saisie dans le champ de contact du titulaire, contenait un nombre acceptable de chiffres. Il est à noter que cette notation n'a pas été utilisée pour les enregistrements des titulaires mais seulement pour les enregistrements techniques ou administratifs.

Un enregistrement a reçu une notation « erreur totale (-2) » s'il manquait le numéro de téléphone ou si le numéro de téléphone contenait le mauvais nombre de chiffres en fonction de son code géographique (ce critère a été appliqué aux enregistrements qui avaient un code géographique répertorié et aux enregistrements auxquels s'était ajouté le code géographique grâce à l'adresse).

Description	Explication	Notation NORC si le code d'erreur s'applique
Pas d'erreur	Le numéro de téléphone est présent, contient un code géographique et un nombre acceptable de chiffres en vertu des exigences applicables aux numéros de téléphone du pays.	+2
Erreur minimale	Vu que la mise en forme n'est pas obligatoire pour les numéros de téléphone en vertu de l'ITU-T E.164, les enregistrements avec des problèmes de mise en forme ont reçu la même notation que les enregistrements notés « pas d'erreur (2) ».	+2
Erreur limitée	Le numéro de téléphone manque de code géographique mais le champ d'adresse correspondant note le pays. Le numéro de téléphone, auquel a été ajouté le code géographique en fonction de l'adresse correspondante, contient un nombre acceptable de chiffres en vertu des exigences applicables aux numéros de téléphone du pays.	+0
Erreur majeure	Le numéro de téléphone du contact administratif ou technique ne contient pas le code géographique mais le champ d'adresse de contact du titulaire contient un pays. Le numéro de téléphone, auquel a été ajouté le code géographique du titulaire en fonction de l'adresse, contient un nombre acceptable de chiffres en vertu des exigences applicables aux numéros de téléphone du pays. Remarquez que cette notation n'est pas utilisée pour les enregistrements du titulaire	-1
Erreur totale	L'enregistrement ne contient pas de numéro de téléphone ou le numéro de téléphone contient un nombre incorrect de chiffres en fonction de son code géographique (cela inclut les enregistrements auxquels le code géographique a été ajouté en fonction de l'adresse).	-2

Validation opérationnelle du téléphone

Aux fins de l'étude pilote, DigiCert a défini un numéro de téléphone comme valide sur le plan opérationnel si le numéro n'était pas déconnecté, si la connexion était effectuée dans les 30 secondes après l'avoir composé et s'il n'y avait aucun message d'erreur une fois connecté. En outre, si le numéro composé était occupé, répondait par un service automatisé ou répondait en indiquant que le numéro fonctionnait correctement, le numéro de téléphone a été considéré valide du point de vue opérationnel.

DigiCert a mené ce processus en demandant à son personnel de soutien de composer manuellement chaque numéro. Afin d'assurer un taux plus élevé d'exactitude dans ses efforts, DigiCert a employé du personnel multilingue afin que les messages d'erreur dans des langues autres que l'anglais puissent être traduits.³

Un enregistrement a reçu la notation « pas d'erreur (+ 2) » si le numéro de téléphone répertorié dans l'enregistrement WHOIS pouvait être rejoint correctement en vertu des spécifications susmentionnées.

Aucun enregistrement n'a reçu la notation « erreur minimale », car le numéro soit se connectait, soit ne se connectait pas. En conséquence, DigiCert n'a pas appliqué cette notation aux enregistrements.

Un enregistrement a reçu la notation « erreur limitée (+ 0) » si le numéro de téléphone inscrit dans l'enregistrement WHOIS n'a pas pu être rejoint mais le numéro de téléphone se connectait après avoir ajouté un code géographique correspondant aux informations du pays de l'adresse.

Un enregistrement a reçu la notation « erreur majeure (-1) » si le numéro de téléphone du contact administratif ou technique inscrit dans l'enregistrement WHOIS n'a pas pu être rejoint mais le numéro de téléphone se connectait après avoir ajouté un code géographique correspondant aux informations du pays de l'adresse du titulaire.

Un enregistrement a reçu une notation « erreur totale (-2) » si le numéro de téléphone était manquant ou ne se connectait pas après avoir ajouté des informations du pays de l'adresse.

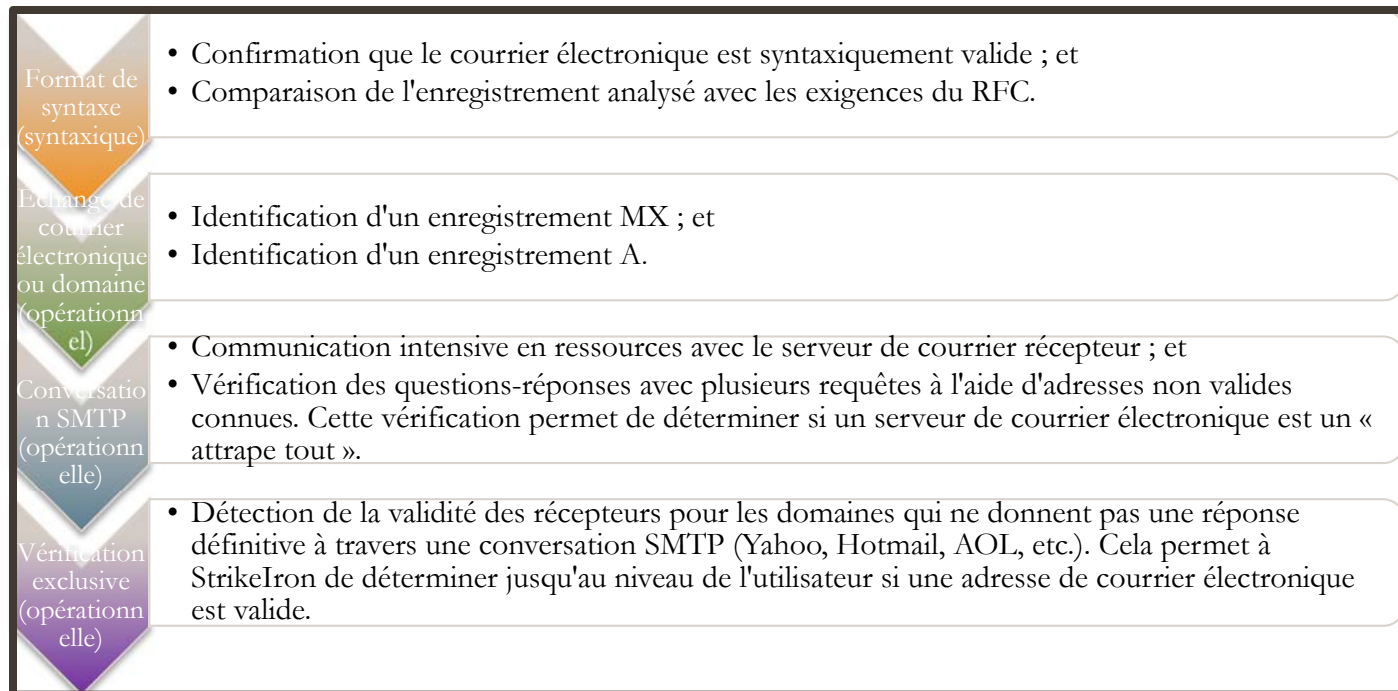
³ Le personnel de soutien de DigiCert parlait les langues suivantes : chinois, anglais, français, allemand, japonais, coréen, portugais, russe et espagnol.

Description	Explication	Notation NORC si le code d'erreur s'applique
Pas d'erreur	Le numéro de téléphone figurant dans l'enregistrement WHOIS se connecte correctement.	+2
Erreur minimale	Aux fins du projet pilote, cette notation n'a pas été incluse.	N/D
Erreur limitée	Le numéro de téléphone inscrit dans l'enregistrement WHOIS ne peut pas être rejoint mais le numéro de téléphone se connecte après ajouter un code géographique correspondant aux informations du pays de l'adresse.	+0
Erreur majeure	Un numéro de téléphone du contact administratif ou technique inscrit dans l'enregistrement WHOIS ne peut pas être rejoint mais le numéro de téléphone se connecte après avoir ajouté un code géographique correspondant aux informations du pays de l'adresse du titulaire.	-1
Erreur totale	Le numéro de téléphone est manquant ou ne se connecte pas après avoir ajouté l'information du pays de l'adresse.	-2

Validation du courrier électronique : Strikelron

L'ICANN a passé un contrat avec StrikeIron pour effectuer l'évaluation de l'exactitude de l'échantillon des adresses de courrier électronique des enregistrements WHOIS.

Le tableau suivant décrit brièvement les mesures générales prises par StrikeIron, avec une explication détaillée dessous.



Validation syntaxique du courrier électronique

Une fois que les enregistrements analysés ont été reçus, StrikeIron a entrepris sa validation syntaxique des adresses de courrier électronique en vérifiant que les adresses de courrier électronique avaient le format syntaxique correct. La norme utilisée pour cette vérification était la conformité avec RFC (p. ex. [RFC 5322](#)).

Plus précisément, StrikeIron a vérifié les courriers électroniques pour s'assurer qu'ils respectaient les quatre spécifications clés de RFC suivantes :

- (1) l'adresse de courrier électronique contenait un symbole « @ » ;
- (2) la partie du domaine de l'adresse de courrier électronique était valide ;
- (3) la partie locale de l'adresse de courrier électronique était valide ;
- (4) l'adresse de courrier électronique respectait toutes les autres exigences du RFC.

Une fois cette étape terminée, StrikeIron a appliqué la notation qui suit aux résultats.

Un enregistrement a reçu une notation « pas d'erreur (+ 2) » s'il se conformait à toutes les spécifications de RFC.

Aucun enregistrement n'a reçu la notation « erreur minimale » car une adresse de courrier électronique est conforme au RFC ou elle ne l'est pas.

Aucun enregistrement n'a reçu la notation « erreur limitée » car une adresse de courrier électronique est conforme au RFC ou elle ne l'est pas.

Aucun enregistrement n'a reçu la notation « erreur majeure » car une adresse de courrier électronique est conforme au RFC ou elle ne l'est pas.

Un enregistrement recevait la notation « erreur totale (+ -2) » s'il ne se conformait pas à toutes les spécifications de RFC.

Description	Explication	Notation NORC si le code d'erreur s'applique
Pas d'erreur	<ul style="list-style-type: none"> • L'enregistrement analysé contient le symbole « @ » ; • L'enregistrement analysé contient un domaine valide ; • L'enregistrement analysé contient une partie locale valide avant le symbole @ ; • La syntaxe de l'enregistrement analysé est valide ; et • L'enregistrement analysé respecte toutes les autres spécifications du RFC. 	+2
Erreur minimale	Aucun enregistrement n'a reçu la notation « erreur minimale » car une adresse de courrier électronique est conforme au RFC ou elle ne l'est pas.	N/D
Erreur limitée	Aucun enregistrement n'a reçu la notation « erreur limitée » car une adresse de courrier électronique est conforme au RFC ou elle ne l'est pas.	N/D
Erreur majeure	Aucun enregistrement n'a reçu la notation « erreur majeure » car une adresse de courrier électronique est conforme au RFC ou elle ne l'est pas.	N/D
Erreur totale	<ul style="list-style-type: none"> • L'enregistrement analysé ne contenait pas le symbole « @ » ; • L'enregistrement analysé ne contenait pas un domaine valide ; • L'enregistrement analysé ne contenait pas une partie locale valide avant le symbole @ ; • La syntaxe de l'enregistrement analysé était non valide ; et • L'enregistrement analysé ne respectait pas toutes les autres spécifications du RFC. 	-2

Validation opérationnelle du courrier électronique

Strikelron a établi trois étapes pour évaluer la validité opérationnelle des adresses de courrier électronique. La première étape était la vérification de l'échange de courrier électronique ou domaine (MX), la deuxième étape était la conversation SMTP et la troisième étape était la vérification exclusive des résultats. Dans le cadre de cette évaluation, Strikelron n'a pas tenté d'envoyer un courrier électronique à l'adresse dans l'échantillon.

Une fois que les enregistrements analysés ont été reçus, Strikelron a pris les adresses de courrier électronique syntaxiquement valides et a réalisé trois vérifications.

La première vérification était une vérification d'échange de courrier électronique ou domaine (MX). Dans cette étape, Strikelron a vérifié si le serveur de courrier électronique pour un domaine indiquait que l'adresse de courrier électronique était valide et ensuite vérifiée.

Pour valider avec des serveurs « attrape tout », Strikelron testerait par la suite une adresse de courrier électronique non valide connue sur le même serveur pour voir si un message d'erreur était reçu pour l'adresse de courrier électronique non valide. Une fois cette étape terminée, Strikelron a appliqué la notation qui suit aux résultats.

Un enregistrement a reçu une notation « pas d'erreur (+ 2) » s'il y avait un enregistrement MX ou un enregistrement A pour le domaine, si le domaine de l'adresse de courrier électronique avait une entrée DNS pour l'enregistrement A, s'il y avait un serveur d'échange de courrier électronique répertorié et si l'enregistrement A répondait comme un serveur de courrier électronique. Un domaine a également pu recevoir une notation « pas d'erreur » s'il avait un enregistrement MX, un enregistrement A et répondait.

Aucun enregistrement n'a reçu la notation « erreur minimale » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.

Aucun enregistrement n'a reçu la notation « erreur limitée » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.

Aucun enregistrement n'a reçu la notation « erreur majeure » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.

Un enregistrement a reçu une notation « erreur totale (+ -2) » s'il n'y avait aucun enregistrement MX ou enregistrement A pour le domaine et le domaine de l'adresse de courrier électronique n'avait pas d'entrée DNS pour l'enregistrement A, s'il n'y avait pas de serveur d'échange de courrier électronique répertorié et si l'enregistrement A ne répondait pas comme un serveur de courrier électronique. Un domaine a également pu recevoir une notation « erreur totale » s'il avait un enregistrement MX, un enregistrement A et s'il ne répondait pas.

Le deuxième contrôle effectué par Strikelron a été une conversation SMTP avec un compte de serveur de courrier électronique pour déterminer si les serveurs de courrier électronique répondaient avec un message « acceptable pour le serveur ». Une fois cette étape terminée, Strikelron a appliqué la notation qui suit aux résultats.

Un enregistrement a reçu une notation « pas d'erreur (+ 2) » si le nom d'utilisateur ou la boîte à lettres était accepté comme une adresse de courrier électronique valide, si l'adresse de courrier électronique était valide conformément aux spécifications RFC et si l'adresse de courrier électronique n'était pas identifiée comme un piège de spam.

Aucun enregistrement n'a reçu la notation « erreur minimale » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.

Aucun enregistrement n'a reçu la notation « erreur limitée » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.

Aucun enregistrement n'a reçu la notation « erreur majeure » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.

Un enregistrement a reçu une notation « erreur totale (+ -2) » si le nom d'utilisateur ou la boîte à lettres était rejeté comme adresse de courrier électronique valide, l'adresse de courrier électronique était non valide conformément aux spécifications du RFC et l'adresse de courrier électronique était identifiée comme un piège de spam.

Description	Explication	Notation NORC si le code d'erreur s'applique
Pas d'erreur	<ul style="list-style-type: none"> • Domaine trouvé : un enregistrement MX ou un enregistrement A existe pour ce domaine. • Domaine de courrier électronique valide : le domaine de l'adresse de courrier électronique a une entrée DNS pour l'enregistrement A, il existe un serveur d'échange de courrier électronique répertorié et l'enregistrement A répond comme un serveur de courrier électronique. • Le domaine de courrier électronique répond : le domaine de l'adresse de courrier électronique a un enregistrement MX et un enregistrement A et répond. • Le nom d'utilisateur ou la boîte à lettres a été accepté comme une adresse de courrier électronique valide. • L'adresse de courrier électronique est valide conformément aux spécifications du RFC. • L'adresse de courrier électronique n'a pas été identifiée comme un piège de spam. 	+2
Erreur minimale	Aucun enregistrement n'a reçu la notation « erreur minimale » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.	N/D
Erreur limitée	Aucun enregistrement n'a reçu la notation « erreur limitée » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.	N/D
Erreur majeure	Aucun enregistrement n'a reçu la notation « erreur majeure » car une adresse de courrier électronique est opérationnelle ou elle ne l'est pas.	N/D
Erreur totale	<ul style="list-style-type: none"> • Domaine non trouvé : aucun enregistrement MX ou enregistrement A n'existe pour ce domaine. • Domaine de courrier électronique non valide : le domaine de l'adresse de courrier électronique a une entrée DNS pour l'enregistrement A ; cependant, il n'existe pas un serveur d'échange de courrier électronique répertorié et 	-2

l'enregistrement A ne répond pas comme un serveur de courrier électronique.

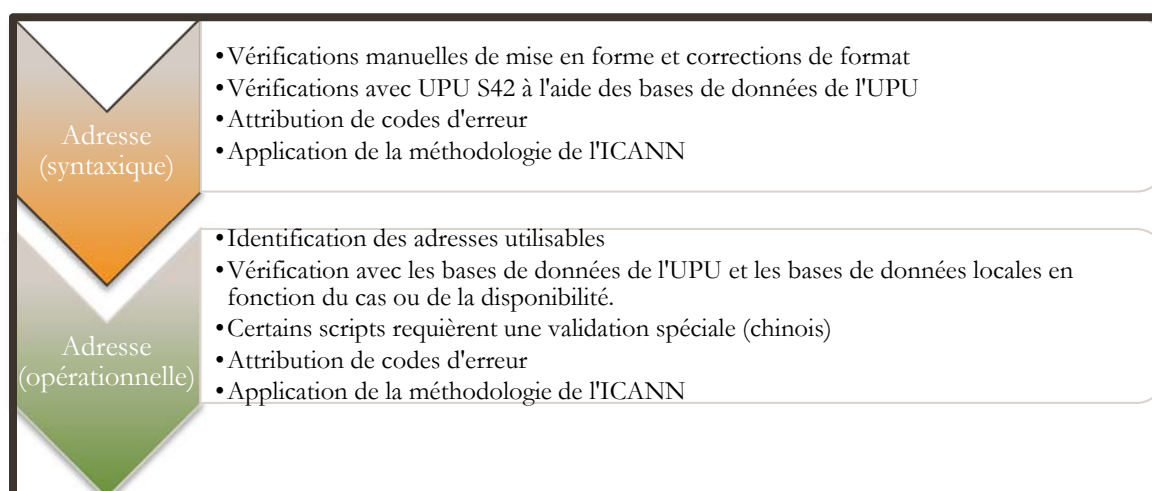
- Le domaine de courrier électronique ne répond pas : le domaine de l'adresse de courrier électronique a un enregistrement MX et un enregistrement A mais ne répond pas.
- Le nom d'utilisateur ou la boîte à lettres a été rejeté comme adresse de courrier électronique valide.
- L'adresse de courrier électronique est valide conformément aux spécifications du RFC.
- L'adresse de courrier électronique a été identifiée comme un piège de spam.

La dernière vérification effectuée par Strikelron a été un test de vérification exclusif pour détecter la validité du récepteur pour les domaines qui ne donnent pas de réponses définitives à travers la conversation SMTP (Yahoo, Hotmail, AOL, etc.). Cette forme définitive de vérification a permis à Strikelron de déterminer, jusqu'au niveau de l'utilisateur, si une adresse de courrier électronique est valide.

Validation postale : Union postale universelle (UPU)

L'ICANN a passé un contrat avec l'Union postale universelle (UPU) pour effectuer l'évaluation de l'exactitude de l'échantillon des adresses postales des enregistrements WHOIS.

Le tableau suivant décrit brièvement les mesures générales prises par l'UPU, avec une explication détaillée ci-dessous.



Validation syntaxique de l'adresse postale

Une fois que les enregistrements analysés ont été reçus, l'UPU a vérifié manuellement que chaque champ d'adresse était dans les champs corrects – si des erreurs étaient trouvées elles ont été corrigées. Par exemple, si un enregistrement analysé contenait toutes les informations d'adresse dans le champ du pays plutôt que dans les champs respectifs, le champ était modifié.

L'UPU a effectué une référence croisée des enregistrements vérifiés manuellement et de la norme d'adressage S42 de l'UPU, qui définit la méthode accordée internationalement pour classer les composantes et les modèles d'adresse postale internationale. Une fois que ces étapes ont été terminées, l'UPU a appliqué la notation qui suit aux résultats.

Un enregistrement a reçu une notation « pas d'erreur (+ 2) » si les composantes de l'adresse sont claires et si les données de l'adresse sont suffisantes pour la validation syntaxique.

Un enregistrement a reçu une notation « erreur minimale (+ 1) » s'il y avait une quantité suffisante de valeurs correctes dans les champs d'adresse pour la validation automatique.

Pour atteindre une notation « erreur limitée (+ 0) », un enregistrement devait avoir au moins une des erreurs suivantes :

- (1) l'enregistrement n'a pas pu être traité, car le pays n'a pas été reconnu ou les champs d'adresse contenaient des valeurs qui n'étaient pas reconnues comme des composantes de l'adresse. Cependant, les données de l'adresse peuvent de toute façon être trouvées exactes lors d'une validation manuelle ultérieure de l'adresse.
- (2) l'enregistrement n'a pas pu être traité, car les données de référence pour ce pays n'étaient pas valides. Cependant, les données de l'adresse peuvent de toute façon être trouvées exactes lors d'une validation manuelle ultérieure de l'adresse.
- (3) certains champs de l'adresse contiennent plusieurs composantes de l'adresse ou plus de données que nécessaire. Cependant, il existe des composantes de l'adresse suffisantes pour livrer à cette adresse et il est possible de la corriger.

Aucun enregistrement n'a reçu la notation « erreur majeure (-1) » car il est toujours possible de corriger les adresses avec plusieurs champs d'adresse.

Aucun enregistrement n'a reçu la notation « erreur totale (-2) » car il est toujours possible de corriger les adresses avec plusieurs champs d'adresse.

Description	Explication	Explication détaillée	Notation NORC si le code d'erreur s'applique
Pas d'erreur	Syntaxiquement correcte	Les composantes de l'adresse (valeurs des champs) sont claires et les données de l'adresse sont suffisantes pour cette validation syntaxique.	2
Erreur minimale		Il existe des valeurs correctes suffisantes dans les champs d'adresse pour d'autres validations opérationnelles automatiques.	1
Erreur limitée	Pas traitée	<ul style="list-style-type: none"> • Pays non reconnu ; • Certains champs d'adresse peuvent contenir des valeurs qui ne sont pas reconnues comme des composantes de l'adresse, le champ de l'adresse du titulaire est vide, utilisation de caractères spéciaux, de diacritiques, adresse de renvoi, ou trop de composantes d'adresse sont présentes dans un champ d'adresse unique ; et • Les données de l'adresse peuvent être considérées exactes pour faire manuellement la validation de l'adresse sur la base des données actuelles de l'adresse, ou lors de l'utilisation des données de contact des informations complémentaires (téléphone, nom du titulaire, organisation du titulaire ou contact technique et/ou administratif du titulaire). 	0

Description	Explication	Explication détaillée	Notation NORC si le code d'erreur s'applique
Erreur limitée	Pas traitée	<ul style="list-style-type: none"> • les données de référence pour ce pays sont non valides ; • Nom possible de l'état/province/territoire dans le champ du pays du titulaire ; et • Les données d'adresse peuvent de toute façon être trouvées exactes lors d'une validation manuelle ultérieure de l'adresse. 	0
Erreur limitée	Questions syntaxiques Improbable de livrer des correspondances	<ul style="list-style-type: none"> • Certains champs d'adresse contiennent plusieurs composantes de l'adresse ou deux fois les données d'adresse postale. • Mais il existe des composantes d'adresse suffisantes pour livrer à cette adresse. • L'adresse peut être corrigée. 	0
Erreur limitée	Questions syntaxiques Assez probable de livrer des correspondances	<ul style="list-style-type: none"> • Certains champs de l'adresse contiennent plusieurs composantes de l'adresse ou deux fois les données d'adresse postale ou le code postal est manquant. • Mais il existe des composantes d'adresse suffisantes pour livrer à cette adresse. • L'adresse peut être corrigée. 	0
Erreur limitée	Questions syntaxiques Très probable de livrer des correspondances	<ul style="list-style-type: none"> • Certains champs de l'adresse contiennent plusieurs composantes de l'adresse ou deux fois les données d'adresse postale ou le code postal est manquant. • Il existe des composantes d'adresse suffisantes pour livrer à cette adresse. • L'adresse peut être corrigée. 	0

Description	Explication	Explication détaillée	Notation NORC si le code d'erreur s'applique
Erreur limitée	Questions syntaxiques L'adresse peut être corrigée.	Il existe des valeurs correctes suffisantes dans les champs d'adresse pour d'autres validations opérationnelles automatiques.	1
Erreur majeure	N/D	Pour la validation syntaxique de l'adresse, la notation -2 ou -1 n'a pas été assignée car, lorsqu'il y a plusieurs champs d'adresse, il est souvent possible de déterminer l'adresse correcte et de la corriger.	N/D
Erreur totale		Pour la validation syntaxique de l'adresse, la notation -2 ou -1 n'a pas été assignée car, lorsqu'il y a plusieurs champs d'adresse, il est souvent possible de déterminer l'adresse correcte et de la corriger.	N/D

Description	Explication	Explication détaillée	Notation NORC si le code d'erreur s'applique
Pas d'erreur	Validée Correcte sur le plan opérationnel	Les composantes de l'adresse (valeurs des champs) sont claires et les données de l'adresse sont suffisantes pour fournir une adresse postale correcte.	2
Erreur minimale	Validée L'adresse a été corrigée	Quelques doutes sur certains champs mais les valeurs correctes dans les champs d'adresse sont suffisantes pour fournir une adresse postale correcte.	1
Erreur limitée	Possibilité de livrer des correspondances L'adresse a été corrigée	Certains champs de l'adresse contiennent plusieurs composantes de l'adresse ou deux fois les données d'adresse postale ou le code postal est manquant. Toutefois, il existe des composantes d'adresse suffisantes pour fournir une adresse postale correcte et y livrer des correspondances.	0
Erreur limitée	Bonne possibilité de livrer des correspondances L'adresse a été corrigée	Certains champs de l'adresse contiennent plusieurs composantes de l'adresse ou deux fois les données d'adresse postale ou le code postal est manquant. Toutefois, il existe des composantes d'adresse suffisantes pour livrer la correspondance correctement à cette adresse.	0
Erreur majeure	Peu de possibilités de livrer des correspondances L'adresse a été corrigée	Certains champs d'adresse contiennent plusieurs composantes de l'adresse ou deux fois les données d'adresse postale. Toutefois, il existe des composantes d'adresse suffisantes pour fournir une adresse postale correcte et peut-être y livrer des correspondances.	-1
Erreur totale	Pas traitée	Cet enregistrement n'a pas pu être traité pour une ou plusieurs des raisons suivantes : <ul style="list-style-type: none"> • Pays non reconnu ; • Certains champs d'adresse peuvent contenir des valeurs qui ne sont pas reconnues comme des composantes de l'adresse, le champ de l'adresse du titulaire est vide, 	-2

		<p>utilisation de caractères spéciaux, de diacritiques, adresse de renvoi, ou trop de composantes d'adresse sont présents dans un champ d'adresse unique ; et/ou</p> <ul style="list-style-type: none"> • Les données d'adresse peuvent toutefois être considérées exactes pour davantage de validation manuelle de l'adresse basée sur les données d'adresse actuelles, ou lors de l'utilisation des données de contact des informations complémentaires (téléphone, nom du titulaire, organisation du titulaire ou contact technique et/ou administratif du titulaire). 	
Erreur totale	Pas traitée	<p>Cet enregistrement n'a pas pu être traité pour une ou plusieurs des raisons suivantes :</p> <ul style="list-style-type: none"> • les données de référence pour ce pays sont non valides ; • nom possible de l'état/province/territoire dans le champ du pays du titulaire ; et/ou • les données d'adresse peuvent de toute façon être trouvées exactes lors d'une validation manuelle ultérieure de l'adresse. 	-2

Validation postale opérationnelle

L'UPU a effectué une évaluation pour la validité opérationnelle des adresses physiques afin de déterminer si une adresse était utilisable grâce à une combinaison de procédures manuelles et automatiques. L'UPU définit utilisables comme, « susceptible d'être mise à profit ». Essentiellement, si l'UPU a décidé, sur la base des différentes normes internationales, qu'une lettre pourrait être envoyée et reçue à l'adresse, cette dernière a été considérée comme utilisable.

Une fois que cette vérification initiale a été terminée, l'UPU a vérifié l'adresse avec ses propres bases de données, ainsi qu'avec les bases de données des pays partenaires, pour déterminer plus profondément si une adresse était utilisable. Dans un cas où l'UPU n'était pas sûre de l'utilisation d'une adresse, l'enregistrement a été transféré au service postal du pays pertinent pour confirmer si l'adresse était utilisable. Par exemple, certains pays exigent que l'UPU envoie des demandes de renseignements concernant la facilité d'utilisation au cas par cas.

En outre, si une adresse était dans un script qui n'est pas pris en charge par les logiciels exclusifs de l'UPU, l'UPU a transmis l'adresse au pays partenaire concerné pour demander de l'aide et ainsi confirmer la possibilité d'utiliser l'adresse. Une fois que ces étapes ont été terminées, l'UPU a appliqué la notation qui suit aux résultats.

Un enregistrement a reçu une notation « pas d'erreur (+ 2) » si les composantes de l'adresse étaient claires et les données de l'adresse étaient suffisantes pour fournir une adresse postale correcte.

Un enregistrement a reçu une notation de « erreur minimale (+1) » s'il y avait des doutes sur certains champs, mais il y avait des valeurs correctes suffisantes dans les champs d'adresse pour fournir une adresse postale correcte.

Un enregistrement a reçu une notation « erreur majeure (-1) » si : (1) certains champs d'adresse contenaient plusieurs composantes d'adresse, deux fois les données d'adresse nécessaires, ou le code postal manquait. Toutefois, il existait des composantes d'adresse suffisantes pour fournir une adresse postale correcte qui pourraient la rendre utilisable.

Un enregistrement a reçu une notation « erreur majeure (-1) » si certains champs d'adresse contenaient plusieurs composantes de l'adresse ou deux fois les données d'adresse. Toutefois, il existait des composantes d'adresse suffisantes pour fournir une adresse postale correcte qui pourraient la rendre utilisable.

Un enregistrement a reçu une notation « erreur totale (-2) » si l'enregistrement n'a pas pu être traité pour une ou plusieurs des raisons suivantes : le pays n'a pas été reconnu, certains champs d'adresse contenaient des valeurs qui ne sont pas reconnues en tant que composantes de l'adresse, le champ était vide ou les données de référence pour ce pays ne sont pas valides. Cependant, les données de l'adresse peuvent de toute façon être trouvées exactes lors d'une validation manuelle ultérieure de l'adresse.