

At-Large Translation and Transliteration of Contact Information PDP Initial Report Workspace

Comment Close Date	Statement Name	Status	Assignee (s)	Call for Comments Open	Call for Comments Close	Vote Announcement	Vote Open	Vote Reminder	Vote Close	Date of Submission	Staff Contact and Email
11.01.2015	Translation and Transliteration of Contact Information PDP Initial Report	ADOPTED 13Y, 0N, 0A	Satish Babu	14.01.2015	20.01.2015 23:59 UTC	23.01.2015 23:59 UTC	23.01.2015 23:59 UTC	29.01.2015	30.01.2015 23:59 UTC	31.01.2015	Julie Hedlund Lars Hoffmann Pol icity-staff@icann.org

For information about this PC, please click [here](#) >>

- [Comments Forum](#)

Brief Overview

Obtain input on the recommendations put forward in this Initial Report by the members of the Translation and Transliteration of Contact Information PDP Working Group. The Group specifically encourages the submission of arguments supporting/opposing its decision not to recommend mandatory translation/transliteration of contact information in this Initial Report.

Comment Period: 16 Dec 2014 23:59 UTC - 11 Jan 2015 23:59 UTC

Reply Period: 12 Jan 2015 23:59 UTC - 1 Feb 2015 23:59 UTC

Section I: Description, Explanation, and Purpose

In addition to background information, an overview of the Working Groups's deliberations and community input received to date, the Initial Report contains the following preliminary recommendations:

Deliberation on Charter Question 1

Is it desirable to translate contact information to a single common language or transliterate contact information to a single common script?

A key issue that emerged early on in the Group's discussion was the agreement that their recommendation should bear in mind that the main purpose of transformed data is to allow those not familiar with the original script of a contact information entry, to contact the registrant. This means that the accuracy of contact information data that are entered and displayed is paramount. There remains however some divergence in the Working Group about whether the need for accuracy is an argument in favour of transformation or not – and this is also reflected in the public comments received (see 'Community Input' below).

At this stage, the Working Group has decided to summarise its discussion and put the arguments it has gathered to the community. The summary provide both detailed arguments in favour and opposing mandatory transformation and the Working Group hopes that community feedback will maximise its consensus level for the Final Report. Therefore, Working Group members **strongly encourage the Community to provide additional arguments in favour/opposing mandatory transformation of contact information data** further to facilitate the WG's consensus building process.

Working Group's arguments supporting mandatory transformation of contact information in all generic top-level domains

- Mandatory transformation of all contact information into a single script would allow for a transparent, accessible and, arguably, more easily searchable database. Currently all data returned from the Whois database in generic top level domains (gTLDs) are provided in ASCII and such uniformity renders it a very useful global resource. Having a database with a potentially unlimited number of scripts /languages might create logistical problems in the long run.
- Transformation would to some extent facilitate communication among stakeholders not sharing the same language. Good communication inspires confidence in the Internet and makes bad practices more difficult. At this stage ASCII/English are the most common script /language choices. However, it should be noted that already today many users of the Internet do not share English as a common language or the Latin script as a common script. The number of these users will grow substantially as internet access and use is continues to expand across countries/continents and so the dominant use of English might deter participation of those not confident in or familiar with it.
- For law enforcement purposes, when Whois results are compared and cross-referenced, it may be easier to ascertain whether the same registrant is the domain holder for different names if the contact information are transformed according to standards.
- Mandatory transformation would avoid possible flight by bad actors to the least translatable languages³.

Working Group's arguments opposing mandatory transformation of contact information in all generic top-level domains

- Accurate transformation is very expensive and these recommendations could effectively shift the costs from those requiring the work to registrars, registrants or other parties. Costs would make things disproportionately difficult for small players. Existing automated systems for transformation are inadequate. They do not provide results of sufficient quality for purposes requiring accuracy and cover fewer than 100 languages. Developing systems for languages not covered by transformation tools is slow and expensive, especially in the case of translation tools. For purposes for which accuracy is important, transformation work often needs to be done manually.⁴ For example the translated 'Bangkok' is more useful internationally than the transliterated 'krung thep'. However, the transliterated 'beijing' is much more useful than the translated 'Northern Capital'. Automated systems would not be able to know when to translate and when to transliterate.
- Another consequence of the financial burden of transforming contact information data would be that the expansion of the Internet and provision of its benefits became more difficult, especially in less developed regions that are already lagging behind in terms of internet access and often don't use Latin-based scripts.
- It would be near impossible to achieve high levels of accuracy in transforming a very large number of scripts and languages – mostly of proper nouns – into a common script and language. For some languages standards do not exist; for those where there are standards, there may be more than one, for example, for Mandarin, Pinyin and Wade Giles.
- Mandatory transformation would require validation of both the original and transformed contact information every time they change, a potentially costly duplication of effort. Responsibility for accuracy would rest on registrants who may not be qualified to check it. Consistent transformation of contact information data across millions of entries is very difficult to achieve, especially because of the continued globalisation of the Internet with an increase in users whose languages are not based on the Latin script. A Domain Name Relay Daemon should display what the client enters. Original data should be authoritative, verified and validated. Interpretation and transformation may add errors.
- Mandatory transformation into one script could be problematic for or unfair to all those interested parties that do not speak/read/understand that one script. For example, whereas transformation from Mandarin script to a Latin script might be useful to, for example, law enforcement in countries that use Latin scripts, it would be ineffectual to law enforcement in other countries that do not read that Latin script.
- A growing number of registered name holders do not use Latin script, meaning that they would not be able to transform their contact information themselves. Therefore, transformation would have to take place at a later stage, through the registrar or the registry. Considering the number of domain names in all gTLDs this would lead to considerable costs not justified by benefits to others and be detrimental to accuracy⁵ and consistency – key factors for collecting registered name holders' contact information data in the first place.
- The usability of transformed data is questionable because registered name holders unfamiliar with Latin script would not be able to communicate in Latin script, even if their contact information was transformed and thus accessible to those using Latin script.

It would be more convenient to allow registration information data to be entered by the registered domain holders in their local script and the relevant data fields to be transformed⁶ into Latin script by either the registrar or the registry. This would provide greater accuracy than transformation and it would provide those wishing to contact name holders to identify their email and/or postal address. A similar method is already in place for some of the country code top level domains (ccTLDs)

Deliberation on Charter Question 2

Who should decide who should bear the burden [of] translating contact information to a single common language or transliterating contact information to a single common script?

The Working Group spent most of its time debating the first Charter question as the answer to this second is very much dependent on the outcome of the first. At this stage, the Group believes that if mandatory translation and/or transliteration would be recommended, the burden of translation/transliteration will probably fall to the operating Registrars who would be likely to pass on these additional costs to their registrants. As stated below, the Working Group would encourage the Community to voice its views on this issue; this includes

Preliminary Recommendations

Preliminary Recommendation #1 The Working Group could recommend that it is not desirable to make transformation of contact information mandatory. Any parties requiring transformation are free to do it ad hoc outside the Domain Name Relay Daemon.

Preliminary Recommendation #2 The Working Group could recommend that any new Registration Directory Service (RDS) databases contemplated by ICANN should be capable of receiving input in the form of non-Latin script contact information. However, all data fields of such a new database should be tagged in ASCII to allow easy identification of what the different data entries represent and what language/script has been used by the registered name holder.

Preliminary Recommendation #3 The Working Group could recommend that registered name holders enter their contact information data in the language or script appropriate for the language that the registrar operates in.

Preliminary Recommendation #4 The Working Group could recommend that the registrar and registry assure that the data fields are consistent, that the entered contact information data are verified (in accordance with the Registrar Accreditation Agreement (RAA)) and that the data fields are correctly tagged to facilitate transformation if it is ever needed.

Preliminary Recommendation #5 The Working Group could recommend that if registrars wish to perform transformation of contact information, these data should be presented as additional fields (in addition to the local script provided by the registrant), to allow for maximum accuracy.

Preliminary Recommendation #6 The Working Group could recommend that the field names of the Domain Name Relay Daemon be translated into as many languages as possible.

"Non-Recommendation" #7 Based on the recommendation #1-#6, the question of who should bear the burden translating or transliterating contact information to a single common script is moot.

*Note: The Working Group in its discussions so far pointed out that regardless of who decides, it is most likely registrars and registrants that would have to carry the financial burden. The Community is **strongly encouraged to supply its views on this issue, regardless of whether they view mandatory translation/transliteration as recommendable.***

¹ 'Transformed' is used throughout this Report, meaning 'translated and/or transliterated'; similarly 'transformation' is to mean 'translation and/or transliteration'.

The AGB defines "searchable" on p.113:

A Searchable Whois service: Whois service includes web-based search capabilities by domain name, registrant name, postal address, contact names, registrar IDs, and InternetProtocol addresses without arbitrary limit. Boolean search capabilities may be offered. The service shall include appropriate precautions to avoid abuse of this feature (e.g., limiting access to legitimate authorized users), and the application demonstrates compliance with any applicable privacy laws or policies.

3 However, it should be noted that transformation tools may not exist for such languages and so transformation would need to be manual until they did. It would be difficult to limit languages to e.g. only the UN ones or some other subset.

4 See: *Study to evaluate available solutions for the submission and display of internationalized contact data* for further information <https://www.icann.org/en/system/files/files/transform-dnrd-02jun14-en.pdf> [PDF, 987 KB].

5 "Accuracy" as used in the "Study to Evaluate Available Solutions for the Submission and Display of Internationalized Contact Data" June 2, 2014:

"There are at least three kinds of use the transformed contact data in the DNDR may have in another language or script (based on the level of accuracy of the transformation):

1. Requiring accurate transformation (e.g. valid in a court of law, matching information in a passport, matching information in legal incorporation, etc.)
2. Requiring consistent transformation (allowing use of such information to match other information provided in another context, e.g. to match address information of a registrant on a Google map, etc.)
3. Requiring ad hoc transformation (allowing informal or casual version of the information in another language to provide more general accessibility)"

Both accuracy and consistency would suffer if large number of actors, for example, registrants, were transforming contact information.

6 "Transformation" on its own is used to mean to refer to contact information, not fields, in this report. A future system could provide field names in the six UN languages and a consistent central depository of field names in additional languages for those registrars et al. that require them for display for various markets.

Section II: Background

The Translation and Transliteration of Contact Information Policy Development Process (PDP) Working Group is concerned with the way that contact information data – commonly referred to as 'Whois' – are collected and displayed within generic top-level domains (gTLDs). According to the [Charter](#) [PDF, 185 KB] (see also Annex A), the PDP Working Group "is tasked to provide the GNSO Council with a policy recommendation regarding the translation and transliteration of contact information. As part of its deliberations on this issue, the PDP Working Group should, at a minimum, consider the following two Charter questions:

- Whether it is desirable to translate contact information to a single common language or transliterate contact information to a single common script?
- Who should decide who should bear the burden [of] translating contact information to a single common language or transliterating contact information to a single common script?

Section III: Relevant Resources

Section IV: Additional Information

Final Issue Report: <http://gnso.icann.org/en/issues/gtlds/transliteration-contact-final-21mar13-en.pdf> [PDF, 654 KB]

Charter: <http://gnso.icann.org/en/issues/gtlds/transliteration-contact-charter-20nov13-en.pdf> [PDF 185 KB]

Section V: Reports

Staff Contact

Julie Hedlund, Lars Hoffmann
Policy-staff@icann.org

FINAL VERSION TO BE SUBMITTED IF RATIFIED

[Please click here to download a copy of the pdf below.](#)

FINAL DRAFT VERSION TO BE VOTED UPON BY THE ALAC

A.1 ALAC commends the Translation and Transliteration of Contact Information PDP Working Group for the work done so far on this important area.

A.2 ALAC considers that Contact Information such as Whois forms an important part of metadata that drives the Internet, and may be of interest to many classes of users including end-users and law-enforcement.

A.3 ALAC feels that transparent access to this information, in a form that can be readily used by end-users, is an important factor that enhances end-user trust and confidence in the Internet.

A.4 ALAC recognizes that in the interests of diversity, Registrants should be allowed to provide Whois information in their own languages and scripts. However, in the interests of the rest of the end-users on the global Internet, ALAC considers that it would be desirable to transliterate and/or translate ('transform') the information provided in non-English languages/scripts to English/Roman.

A.5 If there is no requirement that registration information be also provided in English, it is conceivable that some registrants may attempt to limit transparent access to their Whois information by using relatively obscure scripts and languages for submitting registration information.

A.6 The report points to several potential impediments to making available transformed information, including: (a) difficulty in accurate translations and unavailability of exactly equivalent phrases in English; (b) lack of automated tools for good-quality translations; (c) financial burden of translation; (d) difficulties in re-translating in case of changes to the information provided; (e) the large number of domains, the very large volume of information, and the consequent difficulty in validation of translation; (f) inherent unfairness to information-providers as they may be unable to parse/understand the translated content; and (g) lack of transformation standards that may result in multiple valid but incompatible transformations.

A.7 However, considering increased globalized use of the Internet around the world, it is likely that some of these issues may not pose difficulties in the intermediate term. At this time, English remains the most widespread link language used by diverse communities to understand each other. Many or most registrants are likely to use English as a second language at least, since many technical aspects of the domain hosting require its use (e.g. configuring servers using Linux or Window, programming web sites, accessing technical documentation, using credit cards). Further, automated transformation tools do exist for languages used by some of the very large communities (such as Chinese & Hindi).

A.8 Whereas having registration information in English is desired, several impediments make mandatory translations infeasible (whether by registrants or registrars).

Considering the above, ALAC would like to recommend the following.

Recommendations

B.1 As proposed in the Preliminary Recommendation #1, transformation of contact information does not have to be mandatory. However, there should be a provision for it to be maintained in two forms: a mandatory 'canonical' form in the original language, and an optional 'transformed' form after transliteration/translation. The latter should be a close approximation to the original that can be parsed, understood and used by other communities.

B.2 All ICANN databases, forms and documents should provide for capturing, displaying, storing and maintaining both the forms.

B.3 Registrars should provide Registrants with the option of entering both forms while creating new entries or editing existing entries.

B.4 In the intermediate term, transformation & validation of contact information should be taken up through collaborative efforts of Registrars and the larger ICANN community. In order to minimize costs, such transformation should be done using a combination of automated tools, crowd-sourced community efforts where possible, and encouraging Registrants to enhance their own credibility by providing information in English as well.

FIRST DRAFT SUBMITTED

A.1 ALAC commends the Translation and Transliteration of Contact Information PDP Working Group for the work done so far on this important area.

A.2 ALAC considers that Contact Information such as Whois forms an important part of metadata that drives the Internet, and may be of interest to many classes of users including end-users and law-enforcement.

A.3 ALAC feels that transparent access to this information, in a form that can be readily used by end-users, is an important factor that enhances end-user trust and confidence in the Internet.

A.4 ALAC recognizes that in the interests of diversity, Registrants should be allowed to provide Whois information in their own languages and scripts.

However, in the interests of the rest of the end-users on the global Internet, ALAC considers that it would be desirable to transliterate and/or translate ('transform') the information provided in non-English languages/scripts to English/Roman.

A.5 If there is no requirement that registration information be also provided in English, it is conceivable that some registrants may attempt to limit transparent access to their Whois information by using relatively obscure scripts and languages for submitting registration information.

A.6 The report points to several potential impediments to making available transformed information, including: (a) difficulty in accurate translations and unavailability of exactly equivalent phrases in English; (b) lack of automated tools for good-quality translations; (c) financial burden of translation; (d) difficulties in re-translating in case of changes to the information provided; (e) the large number of domains, the very large volume of information, and the consequent difficulty in validation of translation; (f) inherent unfairness to information-providers as they may be unable to parse/understand the translated content; and (g) lack of transformation standards that may result in multiple valid but incompatible transformations.

A.7 However, considering increased globalized use of the Internet around the world, it is likely that some of these issues may not pose difficulties in the intermediate term. At this time, English remains the most widespread link language used by diverse communities to understand each other. Many or most registrants are likely to use English as a second language at least, since many technical aspects of the domain hosting require its use (eg., configuring servers using Linux or Window, programming web sites, accessing technical documentation, using credit cards). Further, automated transformation tools do exist for languages used by some of the very large communities (such as Chinese & Hindi).

A.8 Given the desirability of having registration information in English on the one hand, and the existence of several impediments that make mandatory translations infeasible (whether by registrants or registrars), ALAC would like to propose the following approach:

B.1 Whois information is to be maintained in two forms: a mandatory 'canonical' form in the original language, and an optional 'transformed' form after transliteration/translation. The latter should be a close approximation to the original that can be parsed, understood and used by other communities.

B.2 All ICANN databases, forms and documents should provide for capturing, storing and maintaining both the forms.

B.3 Registrars should provide Registrants with the option of entering both forms while creating new entries or editing existing entries.

B.4 Over a period of time, the gaps in information or its validation may be taken up through collaborative efforts of Registrars and the larger ICANN community using a combination of automated tools; crowd-sourced community efforts where possible; and encouraging Registrants to enhance their own credibility by providing information in English as well.