This comment is submitted on behalf of the Registrar Stakeholder Group.

Registrars would like to thank the Chair and members of the Policy Working Group for their hard work in preparing this initial report, examining the issue of whether or not to provide mandatory transformation of WHOIS data. It is helpful to have the arguments for and against mandatory transformation set out and explained with such clarity.

Registrars support a more linguistically diverse domain name system, but do not believe that mandating the transformation of WHOIS data is the way to achieve this. The development and operational costs of this obligation would significantly outweigh perceived benefits. It would introduce complexity in maintaining and updating WHOIS records, and in efforts to enforce compliance on the part of service providers.

All this could potentially lead to a registrant losing a domain name for inaccurate data that they did not provide and expressed in a language they do not understand.

**We oppose mandatory transformation of WHOIS data.**

In response to the specific request in the call for comments, we strongly support the Working Group’s Preliminary Recommendation # 1 **not** to mandate the transformation of WHOIS data. We believe the reasons not to mandate are much more thorough and compelling than opposing arguments.

Aside from this point, we have the following comments on the initial report.

**Support for a more linguistically diverse domain name system**

We support measures to increase linguistic diversity within the domain name system. Much remains to be done in this sphere. For example, [analysis](http://www.eurid.eu/files/publ/IDNWorldReport2014_Interactive.pdf) of new gTLD applications not only shows that only 5% were non-Latin script, but that 90% of the new gTLD strings themselves (excluding brands, geographic terms and abbreviations) are meaningful in English.



*Source: World Report on Internationalised Domain Names 2014 (EURid, UNESCO).*

We understand that some services, such as شبكة offer an end-to-end registration service in Arabic, and that currently the only part of the process that requires the registrant to have knowledge of Latin script is the WHOIS data. This is clearly unsatisfactory - registrants should be able to enter contact data in their own language, and to do so will enhance the overall accuracy of the distributed WHOIS database.

Ideally, WHOIS data should be treated similar to the postal addressing system, where transformation into a non-local format or script is strictly optional. Ultimately it is the responsibility of the sender to ensure that the recipient can be reached if a different script is used than the one used locally.  For example, a letter may be sent to Japan using either Japanese or ASCII scripts, but when using ASCII the sender must be sure to use the addressing elements and local formatting recommended by the postal authority for Japan. See <http://www.upu.int/fileadmin/documentsFiles/activities/addressingUnit/jpnEn.pdf>

Introducing a mandatory requirement to transform WHOIS data into one or more commonly used languages would not support the goal of linguistic diversity, but would introduce cost, complexity and risk which outweigh the perceived benefits.

**Costs and benefits - is data transformation a proportionate response to the problem?**

The working group’s charter requests that the following question be considered:

* What exactly the benefits to the community are of translating and/or transliterating contact data, especially in light of the costs that may be connected to translation and/or transliteration?

While the question may be implicitly addressed through the exploration of arguments for and against mandating transformation of data, the report would benefit from some scene-setting to better address the question. It would be helpful to have an evidence base in order to evaluate whether the costs associated with address transformation (were it to be mandated) would be proportionate to the expected benefits.

Here is some data from the registrar community, which may be of assistance:

* Mandatory data transformation into one or more languages would require the addition of several database fields. These fields would need to be added to each registry database, and to be supported by every accredited registrar. On the assumption that the destination writing system for transformed data is likely to be US ASCII, the financial burden is likely to fall most heavily on new market entrant registrars from developing countries or regions described as “underserved” by ICANN.
* The report indicates that the reason for exploring data transformation is to assist the work of brand protection and law enforcement professionals in their research. The proportion of domain names subject to a law enforcement query or brand protection intervention is extremely low, approximately 0.1%. The number of domain names that have been subject to a UDRP or similar intervention is even lower. Why should the cost of data transformation for that small percentage burden the remaining 99.9% of registrations?
* Parallels may be drawn with internationalised domain names, which by their nature support a more linguistically diverse user-base than traditional, ASCII domain names. Research into uptake of internationalised domain names suggests that registrations are intensely localised, and correlate highly to countries and territories in which languages supported by the relevant scripts are spoken. For example, of 900,000 .рф domains, 94% of the registrants are based in Russian Federation, and uptake elsewhere - even former Soviet Union countries - is low. See<http://statdom.ru/tld/%D1%80%D1%84/report/summary/>

In our submission, the cost of data transformation would not be proportionate to the expected benefits.

**Responsibility for WHOIS accuracy and data integrity**

An important principle of the domain name ecosystem is that Registrants provide WHOIS data and they are held responsible for its accuracy. If that data is modified by their Registrar, who assumes responsibility for the accuracy of transformed data? How can Registrants comply with verification request of data they did not provide, and perhaps do not understand?

For data integrity, the WHOIS should display what the registrant provided. Modifications by their registrar (or other third party) is likely to cause more data errors, and less accurate data. A number of questions flow from this:

a. Will there be rules or standards governing translation of non-ASCII characters so that it can be done programmatically? Will a common system be used or are we all just relying on free services like Google Translate?

b. If translation cannot be automated and human judgment is required, who is responsible for doing it?

c. If the registrant is responsible for providing translated data, what if they do not know what it should be?

d. What if a third-party disputes the accuracy of a transliteration?

e. Is the registrant’s consent required before a transliteration is published in the WHOIS and can they withhold consent?

f. What if a registrant wants to change an “approved” transliteration?

g. Is a WHOIS verification required every time one of these transliterated fields are updated?

h. Where does the requirement for data transformation end? Could Chinese law enforcement agents require a contracted party to translate/transliterate existing English contact details into Mandarin? Or, what if the original registration was in a third language/script, for example Russian Cyrillic? Would that translation skip English and go directly to Chinese? What is the service provider supported neither of these languages?

**Compliance**

ICANN should consider the budgetary impact of the costly human resources needed to review translated WHOIS data. It doesn’t make much operational sense, and will likely end up with the registrant paying higher fees for something that they never asked for.

**Comments on the bullet points in favour of mandatory data transformation**

In our opinion, the arguments in favour of mandatory data transformation as set out in the initial report are not convincing, and contain unsupported assertions.

*“Transparent, accessible, and arguably, more searchable database”*

Transparency depends on the viewpoint (and primary language) of the searcher. Only 5% of the world’s population speaks English as their primary language (source: Ethnologue). Yet, the domain name system, and the WHOIS database, are overly dominated by English language and Latin script. For a searcher whose primary language is Chinese, Russian, Arabic, and who is not familiar with Latin script, the WHOIS database is currently far from transparent or accessible.

The argument on searchability is weak. Any string can be searched against and patterns detected, regardless of language or script.

*“Transformation would to some extent facilitate communication among stakeholders not sharing the same language”*

As acknowledged in the paragraph, arguments about facilitating communication cut both ways, and the next billion Internet users (some of whom will be registrants of domain names) will not be as familiar with Latin script as the first billions. In any event, if the registrant is not familiar with Latin script or English language, to the extent that they enter their contact details in another language, it is unlikely that communication through English will be facilitated with that individual. Finally, numerous free and paid services already exist to serve this need, making it unnecessary to burden all registrants with the cost of a pre-translated record.

*“Easier to ascertain whether the same registrant is the domain holder for different names if the contact information is transformed according to standards.”*

This argument assumes that transformation of data will always yield clear and consistent results. This is not the case, as the initial report acknowledges. For example, as the [IRD-WG final report found](http://gnso.icann.org/en/issues/ird/final-report-ird-wg-06mar12-en.pdf) (as quoted in the initial report), "many language translation systems are inexact and cannot be applied repeatedly to translate from one language to another....Translation/transliteration may vary significantly across languages using the same script....Two people may translate/transliterate differently even within a language and the same person may translate/transliterate differently at different times for the same language."

*“Mandatory transformation would avoid possible flight by bad actors to the least translatable languages.”*

First, as shown above, the proportion of “bad actors” affecting the domain name system is extremely low, and this argument does not make the case for why all domain names should be affected by mandatory data transformation. Secondly, the statement contains an assumption that there is an objective standard for “least translatable languages”, whereas the reality is that it will depend on the languages spoken by the registrant and whoever is doing the searching.

In general people tend to register and host locally. This is perhaps a surprising phenomenon given the strength of some registrars internationally. For example, on page 5 at<http://www.eurid.eu/files/publ/IDNWorldReport2014_Interactive.pdf>) there is an analysis of country of hosting for gTLD IDNs plus .eu IDNs. This was done based on the IP ranges associated with the domain names. Countries and regions with strong international registrars (eg North America, UK) don't really show any "winner" script. In contrast, Chinese script, Cyrillic, Han (plus Katakana, Hiragana), Thai, Hangul, Arabic script domains tend to be hosted in countries where associated languages are spoken.

Even within large IDN namespaces which offer multiple scripts (eg .com and .net), registrations cluster strongly around popular scripts. There are very small numbers indeed outside of them.

**Conclusion**

Thank you for the opportunity to comment on the initial report. We support internationalisation of WHOIS data, enabling registrants to enter data in their own language and the language of their registrant customers. We believe that the costs and risks associated with mandating data transformation would far outweigh any perceived benefits. The arguments set out in the initial report in favour of mandatory data transformation are not convincing and should not be adopted. We look forward to continuing to contribute to the policy development work on this issue.