**Translation and Transliteration of Contact Information? Informing the WG Debate**

*(work in process)*

One of the key threshold question the WG is charted to answer is whether or not to translate and/or transliterate Contact Information. To inform the debate on this initial yes/no question Staff has collated some information from related policy and advisory efforts in the wider ICANN community.

One issue related to the question of whether to translate/transliterate Contact Information might be the purpose of Contact Information: **who needs access** to this data and why?

The EWG in its [Initial Report](https://www.icann.org/en/groups/other/gtld-directory-services/initial-report-24jun13-en.pdf) has listed a number of purposes and related use cases and rationales (p.11). It might be useful for the T&T WG to go through these and see whether the Group believes that translation/transliteration of Contact Info might be useful for these different purposes of WHOIS data:

1. Domain Name Control
2. Personal Data Protection
3. Technical Issue Resolution
4. Internet Service Provision
5. Individual Internet Use
6. Business Domain Name Purchase or Sale
7. Domain Name Research
8. Legal Actions
9. Regulatory and Contractual Enforcement
10. Abuse Mitigation
11. Malicious Internet Activities

It might be the case that for some issues translation/transliteration might les useful than for others. The WG could assess these purposes and see whether members believe that translation/transliteration would be helpful for any/all/none of these purposes.

It might also be useful for the Group to assess **ccTLD practices** for the [Final Report of the ICANN IRD WG](http://gnso.icann.org/en/issues/ird/final-report-ird-wg-07may12-en.pdf) contains a summary of queries to ccTLD operators (p.11):

In 2010, ICANN staff informally contacted ccTLD operators whose communities

Staff solicited current WHOIS submission and display practices from 16 ccTLDs by asking the following questions:

1. Does your registry allow users to register domain names using characters from local scripts?

2. Does your registry collect and store registration data in US-ASCII in addition to characters from local scripts?

3. Can users of the web interface choose the display language? What languages does your web interface support?

4. Does your registry provide access to registration information via WHOIS/port 43? Can users choose the display language?

The responses from the 16 ccTLD operators are summarized as follows:

• 10 ccTLDs allow users to register domain names using characters from local scripts;

• 10 of 16 ccTLDs support “English” (US-ASCII) and a local language/script, including

Arabic, Chinese, German, Japanese, Lithuanian, Portuguese, Spanish, Swedish;

• All ccTLDs support WHOIS/Port 43; and

• Character set dependencies affect WHOIS client submission and Display (some uses UTF-8, others UTF-16, or ISO-8859).

The [Final Report of the ICANN IRD WG](http://gnso.icann.org/en/issues/ird/final-report-ird-wg-07may12-en.pdf) provides some interesting points regarding the benefits and the potential problems relating to translating and transliterating Contact Data

**Benefits:**

* Contact Information is less useful to the WHOIS service users who are only familiar with languages that require character set support other than US-ASCII; and this community is growing faster than that capable of using US-ASCII
* Many registrants are monolingual, which is the expectation and motivation behind internationalized domain names[[1]](#footnote-1)
* There are also official and unofficial transliteration standards, most notably those from ISO TC 4627 and the U.S. Library of Congress

**Potential Problem Areas:**

* domain registration data is encoded in US-ASCII. This legacy condition is convenient for WHOIS service users who are sufficiently familiar with languages that can be submitted and displayed in US-ASCII
* Many millions of domain name registrants and Internet users today submit their data in US-ASCII, even though their primary language is written using a different script.
* Using US-ASCII is convenient for registrants, registrars and registries, and the installed base of operational WHOIS services that display US-ASCII.
* Many language translation systems are inexact and cannot be applied repeatedly to translate from one language to another. Thus there will be problems with both consistency and accuracy
* Translation/transliteration may vary significantly across languages using the same script.

**Further Point of consideration from the IDN WG:**

* The desirability to use scripts different than US-ASCII should be balanced against the desired uses of the data. While domain registrants may intend to only use their domain "locally" or interact with people in their native script, the nature of the Internet itself means that any domain provisioned on it is available globally.
* Registrars who accept registration data in local scripts (non US-ASCII) should make registration data in the local script available through DNRD-DS output, and should at least include tags to identify languages and scripts (e.g. RFC 5646).

The GAC notes in its [2007 GAC Principles Regarding gTLD WHOIS Services](file://localhost/Users/Lars/Downloads/WHOIS_principles.pdf):

* gTLD WHOIS services could provide sufficient and accurate data about domain name registrations and registrants subject to national safeguards for individual’s privacy in a manner that (1) supports the stability, reliability, security, and global interoperability of the Internet, from both a technical and public trust perspective; and (2) facilitates continuous, timely and world-wide access.
* Stakeholders should work to improve the accuracy of WHOIS data, and in particular, to reduce the incidence of deliberately false WHOIS data.
1. Also worth considering that monolingual users of WHOIS data will not be able to use translated/transliterated Contact Data (unless it is translated/transliterated into their language) [↑](#footnote-ref-1)