

# Study to Evaluate Available Solutions for the Submission and Display of Internationalized Contact Data

## Introduction

As part of the process to implement the internationalized registration data recommendations of the ICANN WHOIS review team, ICANN commissioned a study to document and evaluate the potential solutions submitting or displaying contact data in non-ASCII (American Standard Code for Information Interchange) character sets. In this wiki we provide some background and outline the requirements for the study.

## Study Areas

- Document the submission practices of internationalized registration data at a representative set of gTLD and ccTLD registries and registrars.
- Document the display practices of internationalized registration data at a representative set of gTLD and ccTLD registries and registrars.
- As electronic merchants and online service providers in other industries often have to accommodate submission or display of their content in multiple languages, investigate and document how other e-merchants or web sites manage internationalized contact data.
- Consider and assess the cost and functionality of commercial, open source, or other known but as yet not widely implemented solutions for 1) transliterating internationalized contact information to US-ASCII, 2) translating internationalized contact information to English, 3) transcribing internationalized contact information to US-ASCII, or 4) a mixture of translation, transliteration and transcription.
- Consider and assess the accuracy implications for transliteration and translation of the internationalized contact data
- Based on practices documented in 3.1 and 3.2 and understanding the issues raised in 3.5 and best practices by other e-merchants in 3.3, what are the common best practices registry/registrar could do to minimize these variations so that translation, transliteration or transcription are done in an un-ambiguous way across all registrars/registries. For example, one such practice could be to have automatic translation + user confirmation /validation, if possible.

The final product is to be in the form of a report. The interim deliverables include 1) a study proposal (along with detailed methodology), 2) preliminary report to be posted for public comment, 3) summary of public comments in the standard ICANN form, and 4) final report after incorporating the community feedback gleaned from the public comments received.

## Researchers

### Marc Blanchet, Guillaume Leclanche, Simon Perreault, Viagénie

Marc Blanchet is President of Viagenie, a consulting and R&D firm in advanced IP networking engineering, with focus on IPv6, VoIP, internationalisation and space networking. Marc has been involved in internationalisation of the Internet, as co-chair of the initial internationalized domain names (idn), vcarddav, precis and iri IETF working groups and co-author of internationalisation protocols (RFC3454, RFC3491). Simon Perreault and Guillaume Leclanche, also from Viagénie, are working with Marc on the project.

### Sarmad Hussain

Dr. Sarmad Hussain is currently a professor of Computer Science and holds the Research Chair on Multilingual Computing at Al-Khawarizmi Institute of Computer Science in Pakistan. He holds a doctoral degree in linguistics and his research is focused on linguistics, localization, language computing standards, speech processing and computational linguistics. He has been developing computing solutions for languages spoken across developing Asia, including standards for Unicode encoding, locale and collation.

## Estimated Timelines

- April 15 2014, interim report posted for public comment.

## Additional Information

See the following attached [terms of reference](#).