ICANN Staff Memorandum

To: The Internationalized Registration Data Working Group (IRD WG)

From: Dave Piscitello, ICANN

Subject: Summary of Internationalized Registration Data submission and display alternatives considered by ICANN Staff and SSAC

As requested, this staff briefing paper summarizes the Internationalized Registration Data submission and display alternatives considered by ICANN Staff and SSAC. The purpose of this paper is to assist the IRD-WG in its study by collecting several considerations and "thought projects" into a single document for WG member consideration. SSAC members and ICANN staff are at the disposal of the WG to elaborate on these considerations.

SSAC SAC037: "Standard" internationalization functionality

SAC037, Display and Usage of Internationalized Registration Data, considers the matter of changing or replacing Whois services to accommodate a global user community who would benefit from using characters from local languages or scripts. Recommendation (3) from SAC037 offers an example of how this might be accomplished:

> "3) ICANN should consider the feasibility of having applications that query Registration data services incorporate "standard" internationalization functionality. For example, an application that binds registered domain names to Registration Data could be required to
> a. accept A-label and U-label domain names as input,
> b. return A-label and U-label domain names as output,
> c. store contact information in XML,
> d. retain Unicode encoding of local characters in stored contact information,
> e. accept Record and Key Type data queries in Unicode, and
> f. return responses to record and key type data queries in Unicode,
> g. return responses including character or word variants bundled with each Registration Data query"

Joint SSAC/ICANN staff: Consideration of study questions

SSAC members and ICANN staff jointly considered the study question "Should the maintenance and display of certain registration data be required in US-ASCII to ensure a common denominator for core information display?" One purpose of having a common denominator for core information display is to facilitate automation of WHOIS queries (within the scope of acceptable use policies). As US-ASCII7 has been used for decades as the de facto character set for automation, it seemed useful to study the extent to which TLD operators supported this character set, even when supporting submission and display using characters from local scripts. Staff informally contacted CCTLD operators whose

communities used languages other than English to collect anecdotal data. Staff solicited current WHOIS submission and display practices from 16 CCTLDs by asking:

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-ASCII7 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?

Staff received responses from 16 CCTLD operators, summarized as follows:

- 10 0f 16 CCTLDs registry allow users to register domain names using characters from local scripts
- 10 of 16 CCTLDs support "English" (US-ASCII7) and a local language/script, including
  - Arabic, Chinese, German, Japanese, Lithuanian, Portuguese, Spanish, Swedish
- All CCTLDs support WHOIS/Port 43
- Character set dependencies affect WHOIS client dependencies submission and display
  - Must support UTF-8, UTF-16, or ISO-8859

Staff discussed a formal survey of the full complement of CCTLD operators. Such a study could be arranged at the request of the IRD-WG.

Staff also considered a second study question, "Are there any general principles that registry operators and registrars could adopt to minimize the 'Babel effect' on Registration Data query services and to ensure some uniformity of information display?" As a starting point, staff considered other forms of communication that can be affected by local language usage and display, e.g., global postal services. The International Address Formatting standard of the Universal Post Union merits consideration. This formatting standard says, "The addressee's address shall written legibly in roman letters and Arabic numerals" and "If other letters and figures are used in the country of destination, it shall be recommended that the address be given also in these letters and numerals." An example from the UPU standard follows:

Example 1: the address of an item posted in Great Britain to the United Arab Emirates will therefore, in principle, be written once in roman characters and once in Arabic characters:

MR. OMAR HUSSAIN
P.O. BOX 111
DUBAI
UNITED ARAB EMIRATES

السيد عمر حسين
ص.ب. 111
دبي
الإمارات العربية المتحدة

In conclusion, we would therefore recommend that:

– you write addresses in roman letters and Arabic numerals; if other letters and numerals are used in the destination country, the address should also be written using these letters and numerals;

– you comply with the recommendations of the postal authority of the country in which you post your items as regards indication of the destination country (element ❶);

– you follow the recommendations of the destination country's postal authority for the remainder of the address (elements ❷);

– you always write the name of the destination country in one of the languages used in the country in which you post your items; if this language is not an internationally known language, you should add the name of the destination country in an internationally known language.

See http://www.upu.int/post_code/en/formatting_an_international_address_en.pdf

Related to the second study question is another question. In particular, in the context of WHOIS, is a "must accept and display US-ASCII7 and may also accept and display <local script>" an appropriate policy for registration data?

Staff also considered a third study question, "What information, and in what languages and scripts, should be permitted when collecting and displaying registration data for a (set of) domain name(s)?" SSAC documents SAC027, SAC033, and SAC037 all encourage ICANN community to consider a successor to the WHOIS protocol that can improve accuracy and availability of domain registration data. Staff and SSAC members are currently studying IRIS/CRISP/DREG standards and prior WHOIS studies to inventory service requirements (a separate GNSO study effort). Before enumerating requirements, however, it may be useful to consider the registration data currently collected to understand what data need to be added to support additional/future services. During the Internationalized Registration Data session at the ICANN Seoul Meeting, Mark Kosters (CTO, ARIN) suggested that staff consider ARIN's work with WHOISRWS, and in particular, to note the use of XML encoding of IP address registration data and the use of HTTP as a delivery protocol for WHOIS. During the Seoul session, participants noted the value (attraction) of having a common data schema and delivery protocol for both domain name and IP network registration data.