Intellectual Property Constituency Comment on the Draft Final Report of the WHOIS Review Team Internationalized Registration Data Expert Working Group

April 21, 2015

The Intellectual Property Constituency (IPC) thanks the members of the WHOIS Review Team Internationalized Registration Data Expert Working Group (IRD Working Group) for their work on the Draft Final Report (Report). See https://www.icann.org/public-comments/ird-draft-final-2015-03-09-en.

All Internet users depend upon accurate, accessible and contactable registration data in the WHOIS database; to advance this goal, the IPC submits these comments on the Report. These comments reflect and build on the IPC's previous public comments on the IRD Working Group's Interim Report, submitted to ICANN on May 27, 2014.   See http://ipconstituency.org/PDFs/IPC_comments_IRD_Interim_Report_052714.pdf.

In its previous comments, the IPC urged the Expert Working Group to select proposal 2, in which the script used for WHOIS data would be the script of the TLD itself (in the case of IDN gTLDs) or else US-ASCII. In the Report, the IRD Working Group chose proposal 1, which allows any free form text to be used for WHOIS data, with the restriction that the scripts of the address should be appropriate for the region in which it is located. This rationale reflects a view that it is of paramount importance that "a registrant must not be required to use any specific language(s) or script(s) beyond what would be used in their [sic] ordinary daily routine," Report, p. 18,[1] and that this objective should trump all other considerations and the interests of WHOIS data users.

The IPC respectfully requests that the IRD Working Group reconsider the selection of proposal 1. While the IPC understands the IRD Working Group's rationale in its choice of free form text, such a selection is problematic to the Internet community as a whole because it:

- Allows a script to be used for WHOIS data that is at odds with other requirements for WHOIS data and with the domain names themselves;
- May be expensive and impractical for many registrars and registries to verify and validate as may (and should) become required of them;
- Focuses solely on the presumed interest of the registrant, without appropriately balancing the registrant's needs with the impact on the main interests of many other users of WHOIS data for its intended purpose – to easily identify, locate and contact domain name registrants – and the widely-acknowledged importance of improving the quality of WHOIS data.

---

[1] Unless otherwise specified, all page references are to the Report.

- Adopting proposal 1 creates the risk that bad faith actors will deliberately obscure ownership information about the domain by selecting a seldom-used language/script with no relation to the registrant or the TLD, a risk that must be weighed against the relatively rare occurrence that the registrant has a legitimate need to display ownership information in a language/script other than those representable in ASCII or in the TLD's script.

Adopting an entirely free form text model as proposed by the IRD Working Group would place WHOIS data at odds with other information concerning domain names that is already required to use US-ASCII. For instance, characters in domain names to the right of the "." in the generic top level domains (other than IDN gTLDs) are already in US-ASCII format. It would be inconsistent and impractical to tell registrants that they may use a script other than US-ASCII for WHOIS data for a domain name that already includes US-ASCII. The use of US-ASCII must already be a part of the "ordinary daily routine" of such registrants to the extent that this routine includes writing, typing, and using the gTLD domain name they have registered.

Furthermore, the Report itself would mandate that US-ASCII be used to furnish several other data elements in WHOIS. This includes the name of the Registrar as it appears in the RAA with ICANN (p. 5);[2] the ISO country code, which exists only in ASCII (p. 23);[3] domain status text values, which under RFC 5731 section 2.3 (as cited on p. 23) must be in US-ASCII; data element tags (p. 19); and the character "Z" in all date and time entries (p. 25). Given these facts, when the domain name itself includes US-ASCII characters, the stated rationale for accepting proposal 1 is outweighed by the interests of WHOIS users, who would benefit more from proposal 2.

The adoption of proposal 1 would place an undue burden on domain name registries and on other aspects of the DNS. As the Report clearly notes, this means that every gTLD registry "must accept every language or script for contact information, i.e., accept whatever the registrar provides." Report, p. 19. This requirement to support virtually every type of character script used throughout the entire world would be burdensome and very expensive for registries.

Further, the IPC respectfully disagrees with the IRD Working Group's statement that the use of US-ASCII in gTLDs that employ that script (as contrasted with IDN gTLDs) would conflict with the objective of internationalizing registration data. To the contrary, requiring a single, known and well-established script such as US-ASCII would actually further the objective of internationalizing data, because it would provide a single standard which the entire world could use for all gTLDs that are written in US-ASCII. The goal of internationalized registration data certainly is compatible with the goal of having consistent and predictable WHOIS data.

Additionally the choice of proposal 1 "requires that a standard list and definition of languages and scripts exist," and one that covers every known language and script that a gTLD registrant might choose to use in supplying WHOIS data (p. 16). Such a comprehensive list and

---

[2] All accredited registrars are listed in the official Internic list with a name presented in US-ASCII. See http://www.internic.net/regist.html. ICANN's instructions for applying for registrar accreditation specifically contemplate that this name may be a "translation or transliteration" of the formal entity name of the applicant. https://www.icann.org/resources/pages/instructions-2014-02-05-en#VI.

[3] See https://www.iso.org/obp/ui#search, as cited in Report p. 23.

definition does not exist today (and may be impossible to maintain if it did). By contrast, if proposal 2 were chosen, lists and definitions do exist for every language and script in which a gTLD can be written today (US-ASCII, and those scripts used in IDN gTLDs approved by ICANN for delegation).

Furthermore, a free-form text model allowing the use of any script would run contrary to the purpose of having WHOIS data and would undermine efforts to improve the quality of that data. The main point of having a WHOIS database is to allow others to quickly ascertain who an owner of a particular domain name is and to communicate with them in short order. Allowing registrants to supply WHOIS data in any script they choose, whether or not that script is reflected in the gTLD itself, will inevitably make it more difficult for others to identify the true registrants using WHOIS data. If use were restricted to the script used in the gTLD (US-ASCII, or the script of IDN gTLDs), users of WHOIS data would more readily identify the registrants and be able to communicate with them. Opening up WHOIS data to an unlimited range of free-form texts makes that quick and easy identification much more difficult.

Finally, the Report itself recognizes that these options will greatly complicate efforts to improve WHOIS accuracy, a critical goal for ICANN.[4] "Taking the language and script information from the registrant requires registrars to be able to detect, validate and verify the script in use. This functionality does not exist in today's registrar customer interface" (p. 16). Perhaps more disturbingly, the Report suggests that "a broad requirement to syntactically validate all postal address contact information may be impractical" under option 1 (p. 17). This directly contradicts ICANN's proposal to assess the validity of WHOIS data across the gTLD space, using syntactic validation as the first tier of assessment.[5] It is of little comfort that "operational validation may [still] be possible," since that level of validation is inherently more expensive and time-consuming.[6]

In sum, this proposal runs counter to numerous critical initiatives while adding complexity and expense and reducing the value and utility of WHOIS information.

If the IRD and the ICANN community insist on the adoption of proposal 1 ("Free form text. The language(s) and script(s) of an address should be appropriate for the region in which it is located."), the IPC suggests the following revision, which may help address some (but certainly not all) of the IPC's concerns:

Proposal 1: Free form text. The language and script of an address **MUST** MATCH THE LANGUAGE AND SCRIPT USED BY THE **OFFICIAL NATIONAL POSTAL AUTHORITY** OF THE COUNTRY IN WHICH THE REGISTRANT IS LOCATED.

---

[4] As recommended by the WHOIS Review Team, "ICANN should … use all means available to progress WHOIS accuracy, including any internationalized WHOIS data, as an organizational objective." https://www.icann.org/en/system/files/files/final-report-11may12-en.pdf, at pp. 12-13.

[5] See generally WHOIS ARS Pilot Study, at http://whois.icann.org/sites/default/files/files/ars-pilot-23dec14-en.pdf.

[6] See WHOIS ARS Pilot Study at 7, 32.

By adopting these changes, proposal 1 becomes more narrowly tailored to the benefits that the IRD cites for adopting proposal 1 over proposal 2.[7] This standard also injects a greater degree of predictability and standardization, which may help address some of the difficulties that free form text presents to registrars, registries, and users of the WHOIS data.

For all these reasons, the IPC respectfully requests that this conclusion of the Report be reconsidered, and that proposal 2 be used instead. Barring that conclusion, we urge the IRD to adopt the above proposed changes to proposal 1.

Thank you for considering our views.

Respectfully Submitted,

Intellectual Property Constituency

---

[7] "Given the purpose of postal information is to be able to contact the entity represented, postal regulations require that the contact information be sufficient to be useful to the local delivery agent." Report, p. 22.