T/T Outstanding Questions: June - July 2017

Question Set 1: On Tagging and "Easy Identification" of Languages/Scripts

Question 1: Recommendation 2 states that a "WHOIS replacement system" should contain data fields that "allow for easy identification…of what languages/scripts have been used by the registered name holder." What does "easy identification" mean? Does this imply that *all* registration data must be tagged with a language/script tag following the adoption of the policy (see questions 1a, 1b, and 1c below)?

Answer (Amr Elsadr, ICANN org and former T/T PDP WG member): The requirement to make languages and scripts identifiable is not in any way associated with whether transformation of contact data takes place (remember that transformation is not mandatory under any circumstance). Obligations come into play when "gTLD-providers" put into place a business model that enables domain name holders to register domain names using registration data in their local languages/scripts (see Recommendation 3: "The language(s) and script(s) supported for registrants to submit their contact information data may be chosen in accordance with gTLD-provider business models").

So, apart from recommendation 3 granting gTLD-providers some flexibility in how they meet their obligations concerning language/script identification, the same recommendation should also be considered an internal node in the decision tree of whether language/script identification is actually an obligation. So, if a gTLD-provider chooses (in accordance with its business model) to allow submission of contact information in local languages/scripts, an obligation to make those languages and scripts identifiable becomes necessary. If the provider chooses not to allow such submissions, no obligation for language and script identification is required in that scenario.

There's an impression that the T/T PDP did not create ANY new obligations on contracted parties. This is only partially accurate. It does create new obligations, but only in the specific circumstances mentioned above.

Answer (Roger Carney, GoDaddy and former T/T PDP WG member): I do not believe Recommendation 2 states, suggests or implies "…that all registration data must be tagged with a language/script tag following the adoption of the policy…". During discussions in the PDP I purposefully avoided (as much as I could) using the term "tagging (tag/tagged)". This term seems to have originated from the IRD WG and I did not think it was useful /appropriate to bring up implementation proposals in the PDP. I think Question 1 as written above is interpreting the text/meaning of Recommendation 2 somewhat different than I did when agreeing with the Recommendation in the PDP. The text in the final report is "*Whilst noting that a Whois replacement system should be capable of receiving input in the form of non-ASCII script contact information, the Working Group recommends its data fields be stored and displayed in a way that allows for easy identification of what the different data entries represent and what language(s)/script(s) have been used by the registered name holder.*" This recommendation does not require or even suggest/recommend that any additional data fields be created. We also need to remember that RDAP does not store nor does it display data, as such I don't believe this recommendation intended RDAP to be considered as a replacement WHOIS system in this context.

Answer (James Galvin, Afilias, T/T PDP WG member, IRD Expert WG chair): Use of the word "tag" and the action "tagging" generally implies a technical interpretation of "easy identification" that includes the use of identifiers specified by RFC 5646 "Tags for Identifying Languages". That is one method of "easy identification" but is not the choice I would support in this IRT. I believe the question of "easy identification" of the script used in the content of data fields is easily answered by recognizing that the Unicode Standard explicitly specifies in which scripts a code point is valid. Given a set of code points the intersection of all valid scripts quickly identifies the script in use. It is language identification that is problematic. Unfortunately, what was not sufficiently considered by either the T/T PDP WG or the IRD Expert WG is that identifying a language requires context and, further, that that context is not generally available. Consider too that for some data elements, e.g., notably a Contact Name, there may be more than one language present in the field content. This latter issue is not addressed at all in the EPP protocol and, since that protocol is in the path from a registrant to a directory service display of data (WHOIS or RDAP), even if the language was known by the registrar there is no way to have this information available for display without technical standards work to "update" EPP. Finally, in consideration of the fact that the language identifier is only needed when a transformation is to take place or is indicated to have taken place, it's presence should only be required in those circumstance. This nuanced interpretation places the burden of "easy identification" of language identifier on the transformation action. Since transformation is explicitly not mandatory the solution for "easy identification" of a language identifier does not need to be a standard and the transformer can perform this task in any way that meets their needs. This nuanced interpretation of "easy identification" may or may not require review outside of the IRT group.

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):

Question 1a: The IRD WG—a Non-Consensus Policy Working Group—recommended that *"Unless explicitly stated otherwise, all data elements should be tagged with the language(s) and script(s) in use, and this information **should** always be available with the data element"*. Does use of "should" instead of "must" in this recommendation indicate that tagging data elements with the language(s) and script(s) in use is not an absolute requirement? Under what circumstances did the IRD WG envision that it may be necessary or desirable to explicitly state otherwise?

Answer (Roger Carney, GoDaddy and former T/T PDP WG member): Judging intent of "should"/"must" is a bit tough for me as I was just a loose observer of this IRD WG. But as this language is from a non-consensus policy working group that the T/T PDP WG reviewed and considered (as asked to do so by the GNSO and Board), it's exact form and probably intent was not carried on by the T/T PDP WG. In comparing T/T Recommendation 2 to this IRD recommendation you can see that they are close in meaning but I think purposefully not exact (i.e. "tagging" and "always be available" were not carried through).

Answer (James Galvin, Afilias, T/T PDP WG member, Chair IRD Expert WG): Taken in context, my personal recollection is that the use of "should" is synonymous with the use of "MUST" in an IETF protocol specification context. However, the statement itself must be evaluated in the context of the full document. Section 2.3 describes the technical considerations that manifest given a mandatory requirement to be present. Section 4.3 defines the verb "to tag" as a requirement for knowing with deterministic certainty the language and script in use, explicitly not suggesting a specific solution. Section 7 recommends next steps given acceptance of the recommendations from the final report, in particular a follow-up effort to review the policy implications of the recommendations. Taken all together, the IRD Expert working group is advice to the community regarding an idealized solution to the issue of "internationalized registration data". It should be understood to describe a goal the community should want to attain eventually, given it is not possible to do it all at once.

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):

Question 1b: If the IRD WG recommendation for tagging data elements with language(s) and script(s) was indeed conditional, was this something considered by the T/T PDP WG while developing recommendations requiring easy identification of language(s)/script(s) used by domain name holders?

Answer (Lars Hoffman, ICANN org, T/T PDP observer): In my recollection, the PDP WG, aware of the IRD formulation, did not want to contradict or change the recommendation of IRD. Their intention, to the best of my memory, was to have all languages 'identified' – there is no two-letter ISO standard for 'suhali', 'urdu', or 'arabic' – so the language field should be tagged, in latin script, by the registrar identifying the language. The idea, if I remember correctly, was that registrants would self-identify their language through a drop-down menu, if they cannot do so (because they can't identify their language in latin script), the registrar would need to identify the script for the RDS record.

Answer (Roger Carney, GoDaddy and former T/T PDP WG member): I don't recall if there was specific recognition of "conditionality" of this recommendation in the PDP. As mentioned above and during several prior meetings, I considered all of the IRD recommendations "conditional." The analysis and recommendations from the IRD were very useful but I (and I think many members of the T/T PDP) used it as input not facts or requirements.

Answer (James Galvin, Afilias, T/T PDP WG member, IRD Expert WG Chair): That "tagging" recommended by the IRD WG was not conditional. Although I do not recall explicit discussion of this specific issue in the T/T WG, I know that the IRD final report was considered input and guidance within the T/T WG, which then made appropriate policy recommendations based on considering the IRD final report advice in the context of the scope of work delegated to the T/T WG. Thus, the short answer to your question is "yes".

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):

Question 1c: Several Recommendations mention the identification of "language(s)/script(s)." What does the "slash" mean? Languages *and* scripts? Languages *or* scripts? Languages *and/or* scripts? Determining the meaning of the "slash" has significant impact on the scope and complexity of obligations needed to implement the policy.

Answer (Lars Hoffman, ICANN org, T/T PDP observer): To the best of my recollection, the reason was that Arabic is a language and a script. My recollection was that, principally, it should be the language that is tagged. Because the script info might not be sufficient if you do not speak the language. – I believe that was mainly a concern of the IPC. Again, to the best of my memory.

Answer (Roger Carney, GoDaddy and former T/T PDP WG member): I think the intent was "whatever is needed" for identification.

Answer (James Galvin, Afilias, T/T PDP WG member, IRD Expert WG Chair): I do not recall specifically how "/" was read out during discussions. Nonetheless, I know from a technical perspective the answer has to be "and". In the general case, it simply is not possible to ensure transformation from one form to another and back to the original without loss of information if both a language and a script tag are not present. It is true that in the majority of cases if only one is specified the other can be inferred, but greater success is present when this is done manually than when this is not programmatically.

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):

Question Set 2: General Implementation Coordination Questions

Question 2: T/T Recommendation #7 states: "These recommendations should be coordinated with other WHOIS modifications where necessary and [be] implemented and/or applied as soon as a WHOIS replacement system that can receive, store and display non-ASCII characters, becomes operational." Does this imply that implementation of the T/T Recommendations is dependent on the implementation of RDAP as a "WHOIS replacement system that can receive, store and display non-ASCII characters"? Does this imply that the implementation of the T/T Recommendations should be coordinated with the Next Generation Registration Directory Service PDP? Specifically, with which "WHOIS modification" efforts should the T/T implementation coordinate and should the T/T implementation be dependent on coordination with these other efforts?

Answer (Lars Hoffman, ICANN org, T/T PDP observer): To the best of my memory, the PDP WG wanted to allow for non-latin scripts to be useable in the WHOIS system as soon as possible. RDAP seemed the most logical vehicle for that. However, the PDP WG did not believe they had the authority in their scope to recommend the adoption of RDAP. So, the consensus was: if RDAP is implemented, that is great and we can proceed with our recommendations. If RDAP does not get implemented then we need to come up with another system that must be able to handle all scripts.

Answer (Roger Carney, GoDaddy and former T/T PDP WG member): When the PDP was working on this language, I recognized that the current WHOIS system, with some modifications, including the RDAP component, may be able to support these recommendations, but not seeing how or when these modifications would be introduced I assumed these recommendations would be used as input into the Next Generation RDS PDP. As everyone has heard me say multiple times I never considered RDAP as a WHOIS replacement system and the text in multiple recommendations does confirm this (i.e. RDAP does not collect, store or display data, it is a communications protocol). More generally and overriding, my intent in the PDP was that these recommendations were holistic in nature. All of these recommendations would be considered together and were only applicable if a registry/registrar chose to translate/transliterate and that there was a system in place for collecting, storing, retrieving and displaying of originally collected data and applicable transformations of that data.

Answer (James Galvin, Afilias, T/T PDP WG Member, IRD Expert WG Chair): It is my recollection that because WHOIS-related activities were still in progress and more were envisioned, it was best to leave the answer to that question to be determined when it was needed. Today, I would say "yes", there is a relationship between implementing these recommendations and the ongoing work in the Next Generation RDS PDP working group. It is our responsibility to identify conflicts and state what can be done now and what should be deferred pending the outcome of that work. There is also the second "WHOIS Review Team" being created; we should consider if there are any conflicts with its expected work product(s).

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):

Question 3: Recommendation 3 has been uncontested within the IRT ("The language(s) and script(s) supported for registrants to submit their contact information data may be chosen in accordance with gTLD-provider business models"). The IRT has noted that, in practice, a number of contracted parties are under the impression that RDS contact information can only be provided in ASCII. Can or should this Recommendation proceed independently of the others to establish a policy around this practice while the Implementation Review Team awaits resolution of the other issues detailed in these questions?

Answer (Lars Hoffman, ICANN org, T/T PDP observer): I am not sure I understand that question that is posed. The PDP WG did not want to force a registrar to offer RDS input in any script, because registrars need to certify the data and thus must have the capacity to read the scripts that registrants use when registering a domain name.

In practical terms, the PDP WG thought that if US-based registrar X wants to sell domain names in Sri Lanka, it can do so offering only Latin-based registration. However, the PDP WG thought, that US-based registrar Y might then seize on that opportunity and offer registration in Sinhalese and Tamil (and hire people that can verify registration data in those languages) – and thus the demand for registrar Y's services would grow much more quickly than for registrar X; leading, over time, to a situation where the market, rather than ICANN policy, determines which scripts are offered to registrants to use to register their domain names.

Answer (Roger Carney, GoDaddy and former T/T PDP WG member): First, I think we should (re)educate contracted parties that US-ASCII is not the only output allowed. The Advisory from 27APR2015 clearly recommends use of US-ASCII but does allow non-US-ASCII if it is encoded in UTF-8 (section I. 3). With this knowledge and as I stated above I think that these recommendations should be treated holistically.

Answer (James Galvin, Afilias, T/T PDP WG member, IRD Expert WG Chair): I agree with Roger. Shorter answer to the question is "yes".

Answer (w/ name and affiliation):

Answer (w/ name and affiliation):