**Informal Review Notes**

**Draft Final Report of the ICANN Internationalized Registration Data Working Group**

**20 September 2011, Version 02**

John C Klensin, 2011-09-22

In the interest of a timely response to the working group, I’m sending these notes rather than a more formally-prepared review. Such a review will be prepared and submitted on request.

I’ve grouped and numbered comments for convenience, but no significance should be read into their order.

I have some issues with the draft document that are basically editorial or small technical corrections in nature; those are being supplied separately.

**General Remarks**

G1. This version of the report is vastly improved over the one I looked at some weeks ago. With a few exceptions, it identifies the issues well and most of the recommendations it does make, even those with which I disagree, are sensible. Substantially of the egregious errors that I identified earlier have been fixed, which is a huge improvement in itself. There are still a number of loose ends, but, even though a few of them are noted below, I assume that the WG has them under control and will get them completed before the draft is forwarded further.

G2. Several of the comments below are ultimately about scope: I haven’t tried to guess whether my issues are with the Board’s conception of the problem, the instructions the Board gave the WG and Staff, your understanding of those instructions, or an issue within the report. From my perspective, the registration data problem is a rather large system of inherently interconnected issues. Carving those issues up and delegate them to separate committees who reach separate conclusions and write separate reports tends to obscure issues rather than providing a sound basis for technical and strategic policy-making. I’ve commented on that issue in this review only when it appears to have gotten in the way of a comprehensive report, but the broader problem is a problem nonetheless.

G3. This report apparently assumes a model of who the “user” of Domain Name Registration Data (DNRD) is, but never identifies or defines who those users are. It then talks primarily about submission of information (e.g., to a registrar and across the registrar-registry interface) and examination of those data by present and would-be registrants. It seems to me that one cannot really determine requirements (and acceptable non-requirements) without an explicit user model. While I have no reason to believe that there is broadly-shared consensus within ICANN about users for DNRD, I note that the original and primary purpose of the WHOIS protocol, dating back even before the DNS itself, was for users and system administrators of hosts on the network to obtain contact information for other hosts and domains in order to work toward resolving problems. While that may no longer be the only important use, it has not disappeared and no report on requirements and options for registration data is complete if it focuses almost completely on the registration process and the way in which DNS labels are obtained. Even if this report is intentionally limited to uses of the data by registrants, registrars, and registries, that should be made explicit enough that other groups can be commissioned to look at other users and uses of the data.

G4. Part of that historical requirement that Whois data, and DNRD data more generally, should be usable to obtain contact information for the “owner” of a domain requires that the actual owner and contact information be obtainable. One aspect of the set of issues that the report identifiers as the translation/transliteration question is that internationalized registration data gives a party that is inclined to try to obscure its identity a new tool: by choosing to place information in an obscure script (e.g., one that many users are unlikely to be able to render properly) or obscure language (one for which translators are unlikely to be readily available to users), one can essentially make otherwise-valid data useless as information outside very restricted communities. At least as important, in a world in which indirect registrations (e.g., hidden, proxy, third-party, and leased registrations) and registrar resellers are considered normal, any rules that are established to keep DNRD data useful and informative as well as simply “available” must also extend to hidden data that can be obtained only by extraordinary means. In other words, if a registrar accepts a third-party registration or agrees to hide registrant information, any rules for data accessibility and usability must also extend to the hidden data themselves. As with the comment above, if the WG chooses to interpret this set of issues to lie outside its scope, they should be identified so that they can be more adequately considered elsewhere.

G5. A key recommendation of the draft report (Recommendation 2 and other text) is that the “translation, transliteration and transcription” issues be the subject of a future report rather than covered in this one. I have some additional comments about that choice below but, at this very general level, I believe that this WG should at least make a recommendation as to whether allocation and delegation of non-country IDN TLDs should be postponed until after that report is available and decisions can be made. Certainly if the issue is considered important, but not one about this WG can reach agreement and make a recommendation, relevant stakeholders in the community will need the opportunity to debate whether the lack of an actionable recommendation should be blocking on actual delegations. Conversely, if the WG were in a position to make an interim recommendation about what should be done until consensus emerges about a long-term recommendation, one that would permit keeping things moving forward.

G6. As another side-effect of several of the issues the report addresses (or, at present, does not address), including those mentioned in G3 and G4 above, it that is seems likely that content and accessibility of DNRD will increasingly become the subject of regulatory efforts in some countries. ICANN and the registrar and registry community can reduce that risk by being aware of the various national sensitivities to both adequate data availability and protection of the privacy of individuals and forming policies about requirements on that basis. While I don’t think this raises any additional specific issues that the report must address, explicitly identifying the topic and risk to encourage its inclusion in trade-off discussions and more general discussion in the community would almost certainly be helpful.

**Issues About Particular Sections or Statements**

I1. The statement about how policies are set for ccTLDs in the Section 2.2 definition of DNRD is correct, but a number of other sections describe ccTLD registries as setting their own policies as if they do so completely unconstrained by national law, regulations, or policies. Those other references and discussions should be more like the one in Section 2.2 (or should be rewritten to refer to a common discussion on the topic).

I2. The “over the web interface” comment under DNRD-DS in Section 2.2 is a little misleading. Whatever can be said for Whois, it is a well-specified and standardized protocol that is specific to these kinds of data. “The web interface” is not standardized and indeed differs significant from one TLD registry to the next and sometimes from one registrar to the next.

I3. Definition of IDNs in Section 2.2. The text reads “...Latin letters with diacritical marks, as required by many European languages”. First of all, there are a rather large number of Latin-script characters that are examples of taking either Basic Latin characters (see comment on the draft) or characters in the ASCII repertoire and adding diacritical marks to them. Very popular examples include specialized ligature forms like æ and ĳ and modified Latin characters such as ø and ç. More important to this discussion, the use of Latin-derived characters is not limited to “European languages”. More or less indigenous languages that are written in Latin script today are found on every continent except Antarctica and most of those languages are in no sense “European”.

I4. The paragraph starting “Collection” in Section 3.3 appears to equate “end user” with “Registrant”. I don’t think that is standard or appropriate terminology (see G3 above). More important, it could suggest to a reader that the WG had an extremely narrow and limited perspective. I think the report is *much* more useful than that conclusion would imply.

I5. It would be good to note that “command line” in Section 3.3 is largely unspecified wrt content and format. There is an open, and complex, question as to whether, despite IETF approval as a replacement for RFC 954, RFC 3912 is a competent and generally-accepted specification of Whois or whether, e.g., the IETF community was asleep at the switch and suffering from the delusion that Whois would rapidly become obsolete and the specification irrelevant. Probably the report should avoid that issue to the degree possible.

I6. Section 3.4 indicates that “all ccTLDs support WHOIS/Port 43. Is this the NVT, telnet without optional negotiation,Whois of RFC 954 or some permissive implementation based on a reading of 3912?

I7. The statement “Character set dependencies affect WHOIS client submission and Display (Must support UTF-8, UTF-16, or ISO-8859)” is unclear. Are all of those CCSs supported, or just one of them per ccTLD (there is a partial answer later in the document, but this still needs clarification). And what does “support… ISO-8859” mean? Support for all the 8859 parts? Some specific one, like 8859-1? Something else?

I8. The statement about monolingual users in Section 4.1 that reads “Therefore, it is unreasonable to assume all of them know or be able to enter the registration data in languages other than their local language” doesn’t follow from the discussion. It would be possible to meet the needs described by an appropriately-designed registrant-registrar interface in the local language that provided translation or mapping to one of a small number of standardized forms/ vocabularies.

I9. In Section 4.2, the report (or at least some members of the WG) recommend that email addresses be permitted in i18n (“EAI”) format. For at least two reasons, this is not advisable. First, there is no standard yet and hence no standard-conforming implementations. While standards might be completed within this calendar year, the new specifications are likely to deploy rapidly in some areas but not in most others. Second, even after deployment begins, there will be large areas of the Internet that will not be able to send or receive messages with internationalized addresses. If email addresses are important contact information, ICANN must insist that all-ASCII addresses be supplied (whether or not non-ASCII ones are supplied as well). Permitting internationalized addresses only would be equivalent to discarding the email address as useful contact and identification information. And, for whatever it is worth, the document cited says nothing about “display” of email addresses. Note that, where the text of the report says “some IRD-WG members recommended...”, the summary table says that Email addresses are “Suitable to Internationalize”.

I10. Under “Registration Status” in Section 4.2, it is entirely reasonable to return EPP status codes and let the retrieving client or operation sort out what to present to the user. However, if that it to be done, the report would be more useful if it provided a normative reference to the standardized list of those codes.

I11. In general, the discussion of the translation and transliteration issues in Section 4.3 covers the material very well. It notes, however, that the WG could not reach consensus on what to do and appears to be recommending another study and WG. See comment G5 above.

I12. If the WG is going to use the term “display specification” (e.g., in Section 4.4 and 4.6), it would be wise to define it and to discuss how ICANN can get a handle on that set of issues. When one is working with client-server protocols in which highly localized user interfaces can be developed (e.g., traditional Port 43 Whois) or with web interfaces in which localizations and local configurations in browsers can override most global specifications about presentation and rendering, it is not clear what “introduce display specifications” actually means. I presume the WG does not propose that ICANN go into the global user interface specification business. But the text that follows in Section 4.4 doesn’t talk about “display specifications” at all. It talks about “data formats” and “data registration schema”, which are not the same thing at all.

I13. It is not clear what recommendation is being made for Section 4.5. “These issues need to be addressed urgently” could be interpreted as indicating that the WG has a plan for updating RFC 3912 and standardizing an i18n model for the Whois protocol. I don’t think that is even plausible, but, if the WG has such a plan (and a theory about how to do it without causing an interoperability disaster), the report should spell it out. If the WG believes that the move toward more internationalized information means that it is time to replace Whois with a different protocol and model (and to get the requirement for Port 43 Whois support out of assorted contracts and agreements, independent of what, if anything, replaces it), then it should say that. Asserting that the issues need to be addressed “urgently” also raises the blocking issue discussed in a different context in G5 above: is this urgent enough that allocation and delegation of non-country IDN TLDs should be deferred until the issues are sorted out? And, if not, why not?

I14. Some of the examples in Appendix B appear to be in need of more discussion, especially given that SSAC is the co-convener of this WG. For example the present discussion of Model 2 includes a comment that some WG members wondered whether the model would “create inaccuracies”. Independent of accidental problems, it should be an obvious question whether some of these models would create attack vectors for those with hostile intent (whether directed against Internet users, the DNS, registries, registrars, or others). Note that there may be an interaction between this topic and the general concerns expressed in G4 above, since some of these models seem to imply that the registries, even the “thick” ones, might not have usable contact or identifying information for anyone but a registrar (and, in the case of resellers, perhaps not that either).