

Internationalized Registration Data Working Group

Meeting Notes

10 May 2010

Attendees: Edmon Chung, Rafik Dammak, Jim Galvin, Jeremy Hitchcock, and Bob Hutchinson; from staff: Julie Hedlund, Dave Piscitello, and Steve Sheng.

Action Items: 1) Staff will outline a Model 4 along the lines suggested by Jim Galvin with assistance from Jim; 2) Staff will develop a draft preliminary approach for Working Group consideration as a basis for a Public Forum in Brussels.

Discussion Summary:

The Working Group members discussed questions circulated by Steve Sheng on the email distribution list concerning how to handle transliteration. In particular, Steve asked whether there were any programming libraries and, if so, then would transliteration be possible if it could be provided at low cost. He also asked whether it would be sufficient for the registrar to provide the registrant's contact information in both local script and transliterated into ASCII. On the list Working Group members noted that there is a Google API and a Microsoft translation utility. Rafik noted that the API depends on the language and he was not sure if it could be implemented in a uniform manner for all languages. Julie asked whether it is feasible to recommend transliteration if there is no uniform solution. Edmon noted that a more important question to address is what will the purpose or use of the transliterated information. He emphasized that if we could determine who will use the information then we could decide how the transliteration could work. Jim added that it would be helpful to know what we are trying to achieve with this recommendation. Ram added that the Working Group could recommend some basic guidelines or best practices with respect to transliteration. Jeremy agreed since he was not sure that transliteration, particularly if it was not uniform, would add much value.

The discussion moved into whether one of the three proposed models could be acceptable. Steve Sheng briefly summarized the three models: 1) US ACSII requirement and registration data in local script; 2) no requirement for US ASCII but requires local script; 3) local script and registrar provides transliteration. Jim Galvin asked whether providing transliteration was only for use by law enforcement and, if so, how frequent would be these requests. Dave noted that if transliteration is optional, rather than required, it could encourage malicious uses. However, he added that someone engaged in anti-phishing might be able to use an online tool for transliteration if there was enough information in the Whois record. Ram noted that if transliteration is required it could be an unfunded mandate if there is no free application and thus the Working Group should consider the economic impact. He added that all transliteration is a best guess. Ram suggested that transliteration should be provided as guidance, not as a mandate, because it is an inherently flawed process. Steve noted that many users would benefit from a transliterated Whois and backward compatibility. Ram noted that Port 43 only supports ASCII.

The Working Group members discussed the idea of a recommendation for backwards compatibility. Edmon noted that with respect to Port 43 the Working Group could consider a format this is backwards compatible – perhaps as a 4th model that requires registrars to provide registrant information in local script and backwards compatible to Port 43 in US-ASCII. This could be SCII and code point numeric format for machine readability. Jim asked again what is the goal of transliteration and what is the audience for the information? He suggested that the Working Group should focus on the representation of the data, use of UTF8 in Whois material with a recommendation that items of general utility and core information should be in UTF8 format. Jim described how SMTP mail was originally ASCII7 and that MIME was developed to accommodate the use of UTF8 encoded characters in both mail headers and message bodies. Dave raised question of whether we might come up with a similar strategy for WHOIS and he indicated "perhaps, but perhaps not on port 43."

Jeremy summarized the discussion thus far: 1) there don't appear to be transliteration tools that are uniform for all languages; 2) transliteration would need to be uniform to be useful; 3) far-flung registrars should be able to choose to represent registrant contact information in their local script, but registrars and registries should have the option to choose different solutions. Ram noted that the Working Group also should consider how urgently the data is needed.