



To: Internet Corporation for Assigned Names and Numbers (ICANN)
From: Messaging Anti-Abuse Working Group (MAAWG)
Date: November 17th, 2011
Subject: MAAWG Comments on Draft Final Report of the Internationalized Registration Data Working Group

To whom it may concern:

Thank you for the opportunity to comment on the ICANN Draft Final Report of the Internationalized Registration Data Working Group (hereafter "the draft report"), as announced at <http://www.icann.org/en/announcements/announcement-3-03oct11-en.htm> .

Introduction

The Messaging Anti-Abuse Working Group (MAAWG) is an international nonprofit, industry-led organization founded to fight online abuse such as phishing, botnets, fraud, spam, viruses and denial-of service attacks that can cause great harm to both individuals and national economies. MAAWG draws technical experts, researchers and policy specialists from a broad base of Internet service providers and network operators representing over one billion mailboxes, and from key technology providers, academia and volume sender organizations. The multi-disciplinary approach at MAAWG (<http://www.maawg.org/>) includes education, advice on public policy and legislation, development of industry best practices, guidance in the development of industry standards and the facilitation of global collaboration.

Many MAAWG members routinely work with domain name registration data (DNRD) for a variety of purposes, including, but not limited to:

- Contacting operations and engineering personnel to resolve technical issues that may arise – issues which, if left unresolved, might impact the deliverability of email or the reachability of network destinations for hundreds of thousands, or even millions, of users.
- Assessing and accumulating reputation data about domains whose registrants may subsequently seek to send mail to our systems and customers, host their domains in our data centers or on our networks, or request organizationally validated SSL certificates. Without public access to accurate and usable domain registration data, new domains are effectively like "someone wearing a ski mask running into a bank on a hot day" and, as a result, will often receive an understandably suspicious (if not outright hostile) reception.
- Both law enforcement officers and private anti-cybercrime investigators use domain registration data to identify the owners of malicious domains that have been registered for spam, malware and other cybercrime-related purposes, or for finding points of contact for innocent third-party domains that have been compromised or have otherwise been abused.
- Once one malicious domain has been identified, domain registration data also helps to identify related domains that may be owned by the same party and that are being misused or are at substantial risk of being misused soon.

MAAWG

Messaging Anti-Abuse Working Group

P.O. Box 29920 ■ San Francisco, CA 94129-0920 ■ www.MAAWG.org ■ info@MAAWG.org

Our ability to continue using domain registration data for these and other important purposes will be jeopardized if domain registration data is inappropriately internationalized or otherwise allowed to become less usable. If that were to occur, MAAWG and its membership would be immensely concerned and hence our response today.

Comments on the ICANN Registration Data (IRD) Working Group Findings

MAAWG generally prefers to offer comments in the same order as the topics are introduced in a report. However, in this case, we are making an exception to highlight Finding 4.5, the most important operational issue raised by the ICANN working group in their report. The other topics follow the report order.

Finding 4.5: Addressing the question "Is the Current WHOIS System Capable of Handling the Query and Display of Internationalized Registration Data?" The working group stated:

The IRD-WG agreed that the current WHOIS system is not capable of handling the query and display of internationalized registration data, but determined that there are some workarounds and local conventions that can permit exceptions. In particular, operators that currently provide DNRD-DS using the WHOIS protocol do not support character sets other than US-ASCII in a standard manner, which could present problems if registrants wish to query DNRD in character sets other than US-ASCII. According to RFC 3912, the WHOIS protocol "lacks [a] mechanism for indicating the character set in use ... This inability to predict or express text encoding has adversely impacted the interoperability and, therefore, usefulness of the WHOIS protocol."

In absence of protocol specification, various registries/registrars have adopted ad hoc solutions to support internationalized DNRD. Continued deployment of ad hoc solutions will inevitably lead to widespread inconsistent user experience and interoperability issues, greatly reducing the utility of DNRD.

Thus, these issues need to be addressed urgently. The IRD-WG encourages the ICANN community to identify, evaluate and adopt an alternative DNRD Access Protocol that would meet the needs of internationalization.

Response: MAAWG agrees with the working group that this is an *extremely* urgent issue.

Discussion: Domain name internationalization work is currently well underway. Given this, we find it particularly disappointing that ICANN is just now addressing the issue that WHOIS, the community's primary source of domain name registration data, is technically ill equipped to accommodate internationalized data. The lateness of ICANN's response to this issue suggests its implicit tolerance of the workarounds and local conventions that have developed regionally. Alternatively, if ICANN has been aware of this issue for some time, why has it failed to undertake the appropriate work to address this problem sooner?

That ICANN is currently promoting the registration of internationalized domains, while aware that WHOIS cannot support this data, is unconscionable. It represents a lack of leadership in an important core area of ICANN's remit. This oversight must be a top priority for immediate attention and expedited remediation.

Finding 4.1: We would next like to comment on the draft report's response to another important question, question 4.1, "Is It Suitable To Internationalize Domain Name Registration Data?" The report appropriately states that:

[...] it is desirable for registrants to be able to submit DNRD in character sets other than ASCII. However, this desirability should be balanced against other uses of the data. While domain registrants may intend to only use their domain "locally" or interact with people in their native script, the nature of the Internet itself means that any domain provisioned on it is available globally.

Response: MAAWG agrees the competing requirements described by the Working Group need to be balanced. However, from the perspective of MAAWG, our primary concern is maintaining the global usability of domain name registration data throughout the Internet. The convenience of an individual domain name registrant constitutes a distinctly lower, secondary priority.

Discussion: Regardless of how DNRD may be extended to support internationalized data, the existing operational usability of DNRD data for broad global audiences **must** be retained. Thus, if ICANN elects to extend DNRD data by allowing entries that cannot be cleanly represented as 7-bit ASCII data, it is critical they use an approach consistent with currently deployed infrastructure and tools. Frankly, we view this as quite a challenge. The vast majority of that infrastructure and those tools will not even be 8-bit clean, to say nothing of often having no ability to render non-Latin character sets for Arabic, Chinese, Greek, Japanese, Korean, Russian, Thai and other languages.

In assessing models that might work in these circumstances, we have been particularly pleased with APNIC region ccTLD-supporting entities (such as JPNIC). These entities offer the ability for WHOIS requesters to supply a flag requesting English data as an alternative to data returned in a local language or character set (e.g., Japanese in the JPNIC case).

We recognize that when confronting a choice between Japanese and English, native French speakers might prefer to see French as yet another alternative language, just as Spanish speakers might prefer to see Spanish as yet another alternative, and so on. While understanding this need, MAAWG believes it is unrealistic to expect DNRD to be made fully multilingual, even in the six official languages of the United Nations (Arabic, Chinese, English, French, Russian and Spanish), much less for all the languages of the world.

Instead, as the International Civil Aviation Organization has done for civil aviation air traffic control, we believe ICANN should standardize on English as a core interoperable and seamlessly available universal language for DNRD, while also allowing other languages where desirable and when they do not cause operational issues. We note this approach is consistent with the Working Group's treatment of data elements such as the Registrar's name, as discussed in section 4.2 of the working group's report:

Sponsoring Registrar (RAA 3.3.1.3): The IRD-WG recommends that this data element should be in ASCII to aid law enforcement and intellectual property investigations, and to the extent possible, make it available in local languages and scripts. It is important to note that ICANN's application for registrar accreditation requires applicants to submit a transliteration of "any legal name, street, electronic or mailing address which is not in Latin characters."

We like that approach, and MAAWG recommends it be used for all DNRD data.

Finding 4.2: "What Data Elements Are Suitable To Be Internationalized?"

The Working Group has offered specific recommendations for a variety of DNRD fields.

Response: MAAWG concurs with, and supports, all but one of working group recommendations.

Discussion: MAAWG endorses the recommendations of the WG for the following data elements:

- *Domain names (RAA 3.3.1.1):* The IRD-WG recommended that WHOIS services should return both A-label and U-label representation for the given IDN domains queried.

- ***Sponsoring Registrar (RAA 3.3.1.3):*** The IRD-WG recommended this data element be in ASCII to aid law enforcement and intellectual property investigations, and to the extent possible, make it available in local languages and scripts. It is important to note that ICANN's application for registrar accreditation requires applicants to submit a transliteration of "any legal name, street, electronic or mailing address which is not in Latin characters."
- ***Nameserver names (RAA 3.3.1.2):*** Currently all nameservers are in US-ASCII. However, with IDNs, it is possible that some registrants will compose nameserver names using IDN labels. Some alternatives exist:
 1. Always display the nameserver name in US-ASCII 7 using the A-label. A supporting argument for this choice is that nameserver name information is generally only of technical interest and should be displayed in the same scheme as it is in the DNS.
 2. Display nameserver names in both A-label and U-label (to the extent such information is available). This is consistent with the recommended treatment of the domain name. The IRD-WG recommended this field should continue to be displayed in A-label and, to the extent possible, be displayed in the corresponding U-label.
- ***Telephone/Fax (RAA 3.3.1.7,8):*** The IRD-WG recommended that the ITU-T Recommendation E.123 could be used to internationalize telephone and fax, specifically using the international notation (+31 42 123 4567).
- ***Email address (RAA 3.3.1.7,8):*** With email internationalization efforts in process at the IETF, IRD-WG members suggested new standards resulting from those efforts could also be applied to the DNRD email address field.
- ***Dates (RAA 3.3.1.4,5):*** This includes creation date, expiration date, and domain update date. The IRD-WG members discussed this issue and proposed ISO 8601-200425 as the standard to be used (e.g 2011-09-19T13:54Z). The ISO standard is also used by ASN.1 and in ICANN draft applicant guidebook WHOIS specifications.

Registration Status: Registrars and registries often provide the status of the domain registration. The IRD-WG identified several alternatives as follows:

1. Return the status in a US-ASCII representation of the registrar's choosing;
2. Publish the exact EPP status code and leave it to the clients to decide whether to localize or not;
3. Identify a more easily understood representation;
4. Publish the easily understood representation in mandatory and local character sets; or
5. Any combination of these approaches.

The IRD-WG members chose option 2, since it gives clients the ability to localize this field. Option 2 is also used in the new gTLD Draft Applicant Guide Book (DAG). The EPP status codes are described in RFCs 5730-5734, RFC 3735, 3915, and 4310.

- Finally, the IRD-WG also recommended that ccTLDs using Extensible Provisioning Protocol (EPP) display it in a similar way.

MAAWG objects to, and does *not* endorse, the recommendation of the working group for Entity Names and Addresses, as it appears to potentially allow these to be entered solely in the user's local language and script. The working group states:

Entity Names and Address (RAA 3.3.1.6,7,8): This includes names and addresses of registrants, administrative contacts, and technical contacts. In line with the recommendations in section 5.1, the IRD-WG agreed that registrants should be able [to] submit their data element in the user's local language and script. Additionally, the IRD-WG discussed whether ~~discussed whether~~ (SIC) it is desirable to adopt a "must be present" representation of contact data, in conjunction with local script support for the convenience of local users.

When we look at the working group's recommendation for 4.3 that follows in the report, we see this issue was a cause for consternation and there was also a lack of consensus among working group participants.

Finding 4.3: This section pertains to the question, "Is It Suitable To Support The Translation Or Transliteration of Entity Name and Contact Information Into a Single Script / Language?" The working group stated in its report:

The IRD-WG examined these models and their impacts on registries, registrars, and end users. In the end, the IRD-WG could not reach a consensus on which model to choose . . .

The IRD-WG would like to recommend an Issue Report on this subject. In the interim, the IRD-WG recommended that the DNRD-DS output should at least include tags to identify languages and scripts (e.g. RFC 5646). Thus, those who need to translate, transliterate or transcribe this information would at least know what language and script the contact information is in.

Response: MAAWG urges that entity name and contact information be made available in English using 7-bit ASCII characters, in addition to any other local language or script that might be used.

We do not object to registrants also providing entity name and contact information in local languages or local scripts, provided those alternative renderings can be suppressed from a pure 7-bit ASCII rendering of the DNRD as a requester may prefer ("JPNIC-style"). MAAWG also does not object to proceeding with an issue report on this topic, or to inclusion of RFC 5646 tags, provided those steps do not interfere with our overall goal of continuing the availability of pure 7-bit ASCII DNRD data access.

Finding 4.4: This question asks, "Is It Suitable To Introduce Display Specifications To Address Internationalized Domain Name Registration Data?" The working group stated in its report:

The IRD-WG agrees that it is suitable to introduce display specifications to address internationalized DNRD. Specifically, while standard formats are defined for domain labels, no standard format is required for elements of a domain name registration record (Registration Data), such as contact information, host names, sponsoring registrar and domain name status.

The IRD-WG concluded that the community would benefit from a standard registration data schema (for example in Extensible Markup Language (XML):

- A formal data schema for registration data (for example in XML) would enable end-user clients to better localize the data label.
- A formal data schema for registration data with language tag information would allow better processing of the data.

- In some cases, it is possible that registrars may allow for multiple languages or scripts in the contact data (for example, e.g. Arabic speaking registrants living in the United States to put their names in Arabic language, but their address in English). If this is needed, the language-tag data needs to be at a data-element level.

Response: MAAWG favors efforts at WHOIS standardization and does not object to the working group's recommendations in this section, although we are concerned about the potential for abuse that may be introduced by data-element-level language and script tagging. Specifically, imagine a registrant who intentionally employs a different language or script for each and every element of their DNRD, not because of exceptional linguistic heterodoxy, but simply to make it harder for anti-abuse personnel to determine the registrant's contact information. We would like to see provisions ensuring this type of abuse is not permitted.

Finding 4.6: This section addresses the question, "Is it Feasible To Introduce Submission and Display Specifications to Address with Internationalized Registration Data?" The working group states:

The IRD-WG agreed that it is feasible to introduce submission and display specifications to address internationalized registration data. In particular, most elements of the DNRD have existing standards that apply to them, and to the extent possible, the IRD-WG recommend those be considered. In particular, the following schemas/templates have been worked on in the past and should be considered:

- EPP RFC 5730-5734
- DREG RFC 3982
- UPU S.42 (address templates)

Response: MAAWG does not object to the further consideration of these IETF and Universal Postal Union standards for address standardization purposes.

5. Recommendations: The working group also made three recommendations as part of their report:

Recommendation 1: ICANN staff should develop, in consultation with the community, a data model for domain registration data. The data model should specify the elements of the registration data, the data flow, and a formal data schema that incorporates the standards that the working group has agreed on for internationalizing various registration data elements. This data model should also include tagging information for language/scripts.

Recommendation 2: The GNSO council and the SSAC should request a common Issue Report on translation and transliteration of contact information. The Issue Report should consider whether it is desirable to translate contact information to a single common language or transliterate contact information to a single common script. It should also consider who should bear the burden and who is in the best position to address these issues. The Issue Report should consider policy questions raised in this document and should also recommend whether to start a policy development process (PDP).

Recommendation 3: ICANN staff should work with the community to identify a DNRD Access Protocol that meets the needs of internationalization, including but not limited to the work products resulting from recommendations 1 and 2, and the requirements enumerated in this report.

Response: Subject to the considerations outlined earlier in these comments – and recognizing there will be opportunity for community participation in the outlined activities with community review of the resulting recommendations – MAAWG does not object to these working group recommendations.

Conclusion

We appreciate this opportunity for MAAWG to comment on ICANN's Draft Final Report of the Internationalized Registration Data Working Group. If you would like us to discuss any of our remarks in more depth, or if you have any questions, please do not hesitate to contact us.

Sincerely,

/signed/

Jerry Upton, Executive Director

Messaging Anti-Abuse Working Group (MAAWG)

jerry.upton@maawg.org

<http://www.maawg.org>

PPC2011-15