Draft Final Report of the ICANN Internationalized Registration Data Working Group

24 August 2011

Version 00

Table of Contents

[editor note: general comment, need to double check every public comment received to make sure important contributions and view points are not missed.]

1. Executive Summary

On 26 June 2009 the Board of Directors of the Internet Corporation for Assigned Names and Numbers (ICANN) approved a resolution (2009.06.26.18) requesting that the Generic Names Supporting Organization (GNSO) and the Security and Stability Advisory Committee (SSAC), in consultation with staff, convene an Internationalized Registration Data Working Group (IRD-WG) composed of individuals with knowledge, expertise, and experience in these areas to study the feasibility and suitability of introducing display specifications to deal with the internationalization of registration data.[1] Subsequently, the SSAC and the GNSO formed the IRD-WG to study the feasibility and suitability of introducing specifications to deal with the submission and display of internationalized registration data.

This final report is IRD-WG's response to the Board request regarding the feasibility and suitability of introduction display specifications to deal with the internationalized registration data.

[editor note: low priority: the last paragraph needs to be rewritten to include findings and recommendations from the IRD-WG, but this can be added at the end.]

2. Introduction

With the increasing use of the Internet in all geographic regions and by diverse linguistic groups, the demand for a multilingual Internet has intensified. To satisfy the demand, many Internet applications are now able to accept and to display characters from a broad range of languages and scripts. In addition, the introduction of internationalized domain names (IDN) at the top level of the Domain Name System (DNS) culminates a global effort to fully internationalize domain names.[2]

[editor note: low priority: address JCK's comments. Members of the broader community have explained to ICANN multiple times that the use of IDNs has little to do with a "multilingual Internet" even though they make a slight contribution to enabling use by people who use different languages and writing systems (see, for example, my comments to the ITU on the subject [Klen06]). While the organization, composition, and presentation of registry databases may actually be more relevant to that topic, tying this work to the "demand for a multilingual Internet" mostly just distracts people from the real issues.]

[1] See ICANN Board Resolutions, 26 June 2009, "Display and Usage of Internationalized Registration Data," <http://www.icann.org/en/minutes/resolutions-26jun09.htm#6>.

[2] By September 2010, 15 new IDN top level domains (TLDs), representing 12 countries/territories, have been added to the root zone. Since all are ccTLDs, ICANN has not yet had to confront the issues of internationalized registration data in the gTLD space.

Access to domain name registration information (often called WHOIS data) is provided by several applications. Accommodating the submission and display of internationalized registration data is seen as an important evolutionary step for WHOIS services. The following statement from the Internet Architecture Board (IAB), found in RFC 4690, summarizes the issues associated with this evolution:

> "In addition to their presence in the DNS, IDNs introduce issues in other contexts in which domain names are used. In particular, the design and content of databases that bind registered names to information about the registrant (commonly described as "WHOIS" databases) will require review and updating. For example, the WHOIS protocol itself [Daigle 2004][3] has no standard capability for handling non-ASCII text: one cannot search consistently for, or report, either a DNS name or contact information that is not in ASCII characters. This may provide some additional impetus for a switch to IRIS [Newton and Sanz 2005a, 2005b] but also raises a number of other questions about what information, and in what languages and scripts, should be included or permitted in such databases."[4]

The SSAC also called attention to these issues in SAC037, *Display and usage of Internationalized Registration Data, Support for Characters from Local Languages or Scripts*.[5] In this report, the SSAC recommended that the ICANN Board of Directors form a working group to study the feasibility and suitability of introducing submission and display specifications to deal with the internationalization of registration data. At the request of the ICANN Board of Directors, the GNSO and the SSAC created the IRD-WG to study this issue.

[Editor note 2: add text about about the structure and organization of this report.]
[Editor note: introduction is largely done besides addressing the previous editor notes.]

2.1 IRD-WG Objectives and Membership

Text of the Board Resolution[6]

"Whereas, ICANN has been working towards the introduction of Internationalized Domain Names (IDN) with the gTLD and ccTLD communities.

---

[3] L. Daigle, "RFC 3912: WHOIS Protocol Specification," Network Working Group, Internet Engineering Task Force, Internet Society, September 2004, <http://www.ietf.org/rfc/rfc3912.txt>.
[4] J. Klensin and P. Fältström, "RFC 4690: Review and Recommendations for Internationalized Domain Names (IDNs)," Network Working Group, Internet Engineering Task Force, Internet Society, September 2006, <http://www.ietf.org/rfc/rfc4690.txt>.
[5] Security and Stability Advisory Committee, "SAC037, Display and usage of Internationalized Registration Data, Support for Characters from Local Languages or Scripts," 21 April 2009, <http://www.icann.org/en/committees/security/sac037.pdf>.
[6] See ICANN Board Resolutions, 26 June 2009, "Display and Usage of Internationalized Registration Data," <http://www.icann.org/en/minutes/resolutions-26jun09.htm#6>

Whereas, support for characters from local languages in domain name registration submission and display is an issue that affects many communities across the GNSO, CCNSO, ALAC and GAC.

Whereas, 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.

Whereas, members of the community with knowledge and expertise in these areas have identified topics of inquiry in the display and usage of internationalized Registration Data, including applications and Internet user experience, data reliability, accuracy and operational issues, and security and standardization issues. See:

- SAC037 "Display and usage of Internationalized Registration Data" (21 April 2009) <http://www.icann.org/committees/security/sac037.pdf>;
- SAC033 "Domain Name Registration Records and Directory Services" (22 July 2008) <http://www.icann.org/committees/security/sac033.pdf>;
- SAC027 "Comment to GNSO regarding WHOIS Studies" (7 February 2008) <http://www.icann.org/committees/security/sac027.pdf>.

Whereas, the Board recognizes that discussion and resolution of these issues would be beneficial to the introduction of Internationalized Domain Names.

Resolved (2009.06.26.18), the Board requests that the GNSO and SSAC, in consultation with staff, convene an Internationalized Registration Data Working Group comprised of individuals with knowledge, expertise, and experience in these areas to study the feasibility and suitability of introducing display specifications to deal with the internationalization of Registration Data.

The Board directs the Internationalized Registration Data Working Group to solicit input from interested constituencies including ccTLD operators and the CCNSO during its discussions to ensure broad community input.

The Board further directs staff to provide a dedicated staff person and additional staff resources as staff determines to facilitate the work of the Internationalized Registration Data Working Group.

The working group interprets the Board's request as two questions:

1) *Suitability*: Is it suitable (or desirable) to support characters from local languages in domain name registration submission and display? A related question: is it suitable to introduce display specifications to deal with the internationalisation of Registration data.

2) *Feasibility*: Is it feasible to introduce submission and display specifications to deal with internationalized registration data?

IRD-WG Membership: Edmon Chung from the Generic Name Supporting Organization (GNSO) and James Galvin from the Security and Stability Advisory Committee (SSAC) co-chaired the working group.[7] The working group members include a total of seventeen participants from gNSO, ccNSO, ALAC and SSAC[8]. There is also some geographic diversity in the membership with participants from China, Tunisia, New Zealand, Russia, and the USA.

[editor note: is it desirable to list all the participants in an Appendix?]

2.2 Terminology

The term "WHOIS" in the ICANN could refer to various components of the WHOIS system. To avoid confusion and bring precision to the discussion, in this report we use the following terms as proposed in [SAC tbd] to better distinguish the components of the WHOIS system:

[editor note: The SSAC advisory on terminology is first published as PowerPoint presentations at the Singapore meeting, there will be a full report from SSAC on this]

1. **Domain Name Registration Data**– refers to the information that registrants provide when registering a domain name and that registrars or registries collect. Some of this information is made available to the public. For interactions between ICANN Accredited Registrars and registrants, the data elements are specified in the Registrar Accreditation Agreement.[9] For country-code top-level Domains (ccTLDs), the operators of these TLDs set policy for the request and display of registration information. In this report, we refer primarily to ***registration data.***

2. **Registration Data Access Protocol** – refers to the elements of a (standard) communications exchange—queries and responses—that make access to registration data possible. For example, the WHOIS protocol (RFC 3912) and HTTP (RFC 2616 and its updates) are commonly used to provide public access to registration data.

3. **Registration Data Directory Service (RDDS)** – refers to the service(s) offered by registries and registrars to provide access to (potentially a subset of) the domain name registration data. ICANN Accredited Registries and Registrars are required by contracts to provide the RDDS services via both port 43 and over the web interface.[10] For country-

[7] The WG would also like to acknolwedge that Jeremy Hitchcock has been a chair of the WG representing SSAC till December 2010.
[8] For a list of the IRD-WG members and the Charter, see the IRD-WG wiki at <https://st.icann.org/int-reg-data-wg/index.cgi?internationalized_registration_data_working_group>.
[9] See Section 3.3 of Registrar Accreditation Agreement (2009). <http://www.icann.org/en/registrars/ra-agreement-21may09-en.htm#3>
[10] See Section 3.3 of Registrar Accreditation Agreement (2009). <http://www.icann.org/en/registrars/ra-agreement-21may09-en.htm#3> and Registry Agreements <http://www.icann.org/en/registries/agreements.htm>

code top-level domains (ccTLDs), the TLD operators determine which service(s) they offer.

To ensure that discussions regarding internationalized registration data take place in a consistent manner, the working group uses the following definition of IDN-related terms. These terms are used in consistency with ICANN's IDN glossary[11] and IETF Internationalisation terminologies as defined in draft-ietf-appsawg3536bis-06[12] and RFC 5890[13].

[Editor note: draft-ietf-appsawg-rfc3536bis-06 is not a RFC yet and should not be considered as a stable reference. Alternatively we could simply cite the original RFC 3536.]

***Internationalized domain names (IDNs)**:* IDNs are domain names that include characters used in local languages scripts that are not written with the twenty-six letters of the basic Latin alphabet. An IDN can contain Latin letters with diacritical marks, as required by many European languages, or may consist of characters from non-Latin scripts such as Arabic or Chinese.

***Internationalized Registration Data (IRD)**:* IRD are domain registration data that have at least one data element that is composed of characters used in a local representation of a language other than (case-insensitive) ASCII letters (a-z), digits (0-9), and the hyphen (-).

ASCII (American Standard Code for Information Interchange): A common numerical code for computers and other devices that work with text. Computers can only understand numbers, so an ASCII code is the numerical representation of a character such as 'a' or '@'. When mentioned in relation to domain names or strings, ASCII refers to the fact that before internationalization only the letters a-z, digits 0-9, and the hyphen "-", were allowed in domain names.

A-label | U-label: A domain name consists of a series of "labels" (separated by "dots"). The ASCII form of an IDN label is termed an "A-label." An A-label conforms to the Letter-Digit-Hyphen (LDH) constraint on labels as defined by the DNS standards. All operations defined in the DNS protocol use A-labels exclusively. The Unicode form, which a user expects to be displayed, is termed a "U-label." A special form of "ASCII compatible encoding" (ACE) is applied to a U-label to produce a corresponding A-label. The transformation is symmetric: one can derive a U-label from an A-label for the purpose of displaying the domain name using characters from a local script so that a user sees a familiar script rather than a less recognizable A-label.

Thin | Thick Registry: A thin registry only includes data sufficient to identify the sponsoring registrar, status of the registration, nameserver, creation, and expiration dates for a domain registration. Registrars maintain the complete set of registration data for those domains they sponsor. .COM and .NET are examples of thin registries. Thick registries maintain the

[11] ICANN, "IDNs Glossary," Retrieved August 10, 2010, <http://www.icann.org/en/topics/idn/idn-glossary.htm>.
[12] P. Hoffman and J. Klensein (2011). Terminology Used in Internationalisation in the IETF. draft-ietf-appsawg-rfc3536bis-06. Available at <http://tools.ietf.org/html/draft-ietf-appsawg-rfc3536bis-06>.
[13]

registrant's contact information and designated administrative and technical contact information, in addition to the sponsoring registrar and registration status information supplied by a thin registry. .INFO and .BIZ are examples of thick registries.

[Editor note: medium priority: the terminology section needs to be expanded for any important terms used in the document.]

3. Background

3.1 What is Domain Registration Data?

As defined earlier, domain registration data refers to the information that registrants provide when registering a domain name and that registrars or registries collect. Some of this information is made available to the public. The Registrar Accreditation Agreement (RAA 3.3.1) specifies the following data elements that must be provided by registrars (via Port 43 and via web-based services) in response to a query:

- The Registered Name;
- The names of the primary nameserver and secondary nameserver(s) for the Registered Name;
- The identity of the Registrar (which may be provided through Registrar's website);
- The original creation date of the registration;
- The expiration date of the registration;
- The name and postal address of the Registered Name Holder;
- The name, postal address, e-mail address, voice telephone number, and (where available) fax number of the technical contact for the Registered Name; and
- The name, postal address, e-mail address, voice telephone number, and (where available) fax number of the administrative contact for the Registered Name.

The IRD-WG notes that some registries are subject (in their registry agreements with ICANN) to slightly different requirements regarding which data must be publicly accessible.[14]

In ccTLDs, there is no such requirement, and each ccTLD sets its own policy regarding what constitutes this data. Usually, a subset of the data above is displayed in ccTLD registration data directory services.

The registration data required by RAA can be broken down into the following categories:

[14] See for example <http://www.icann.org/en/tlds/agreements/biz/appendix-05-08dec06.htm.

- Domain name
- Sponsoring Registrar
- Registration status
- Nameserver information
- Names (e.g. owner, admin, technical contact)
- Postal Addresses (owner, admin, technical contact postal information)
- Phone/fax numbers (e.g. admin, technical, owner phone/fax)
- Dates (e.g. creation date, expiration date, update date)
- Email Addresses (owner, admin, technical contact email)

3.2 Where are different registration data elements collected, stored, and displayed?

Collection: End users submit this information as part of the process of registering a domain name. The sponsoring registrar or their reseller collects the information.

Storage: The sponsoring registrar that receives the domain registration data stores a copy. In addition, the registrar submits a limited subset of the information (domain status, nameserver information) to the registry if the sponsoring registry is a thin registry. If the sponsoring registry is a thick registry, registrars submit a more complete set of the information as required by RAA 3.3.1 to the registry.

Display: End users query the registration data directory service for a domain name, contact information, or nameserver information.

- For thick registries, the query displays the data from the registry's WHOIS service.
- For thin registries, the query displays the data from the registry's WHOIS service. Some WHOIS clients could parse the data and continue to query the registrar's WHOIS service.

3.3 Current Practices by gTLD Registries / registrars and ccTLDs to Support the display of Internationalised Data

- No standards, conventions, or policy requirements exist (some data is in non-US-ASCII)
- Support of ASCII7 CRLF is the only obligation
- Registration Data directory services commonly only support submission and display using ASCII7 (remove)
- Certain registries have developed various ad hoc or TLD-specific conventions for signaling (Jay Daley identified several examples)
- To support internationalized data, many registries has developed specific conventions.
- Conventions that are adopted for WHOIS/43 clients may not be used or useful for web based services (i.e., if the web-based WHOIS service accesses registrar or registry data doesn't "proxy" queries onto a port 43 query…)

[editor note:high priority: this is an important section that needs to be substantially rewritten]

3.4 The WHOIS protocol

The ***WHOIS protocol*** (RFC 3912)[15] describes exchanges of queries and messages between a client and a server over a specific port (port 43).

- The protocol imposes no constraints on the data it transports. The only constraint imposed on query and message formats is that they must be terminated using an ASCII line feed (LF) and carriage return (CR) character sequence.

- According to RFC 3912, "The WHOIS protocol has not been internationalized. The protocol has no 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."[16]

- The WHOIS protocol does not define a structured data schema.

4. International Standards

In this section, we list relevant international standards and standard practices that exist that could be used as display specifications for registration data.

4.1 IETF standards

- IDNA 2008 (RFC 5890 – 5894)
- EPP RFC 5730-5734
- DREG RFC 3982

4.2 UPU standards

- UPU E.123 (for telephone/fax number)
- UPU S.42 (address templates)

4.3 Standards for transliteration and translation

- ISO 9:1995 Cyrillic -> LATIN
- ISO 233:1984 Arabic -> LATIN
- ISO 233-2:1993 Arabic -> LATIN, simplified

[15] L. Daigle, "RFC 3912: WHOIS Protocol Specification," Network Working Group, Internet Enginerring Task Force, The Internet Society, September 2004, <http://www.ietf.org/rfc/rfc3912.txt>.
[16] L. Daigle, "RFC 3912: WHOIS Protocol Specification," Network Working Group, Internet Enginerring Task Force, The Internet Society, September 2004, <http://www.ietf.org/rfc/rfc3912.txt>.

- ISO 259:1984 Hebrew -> LATIN
- ISO 843:1997 Greek -> LATIN
- ISO 3602: 1989 Japanese -> LATIN
- ISO 7098:1991 Chinese -> LATIN
- ISO 9984:1996 Georgian -> LATIN
- ISO 9985:1996 Armenian -> LATIN
- ISO 11940:1998 Thai -> LATIN
- ISO/TR 11941:1996 Korean-> LATIN
- ISO 15919:2001 Denanagari -> LATIN

- (add finding based on these standards)
The IRD-WG members agreed that various elements of registration data could be separately internationalized, in particular:

Domain names (RAA 3.3.1.1): The IRD-WG recommends 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 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 RAA requires applicants to submit a transliteration of "any legal name, street, electronic or mailing address which is not in Latin characters."[17]

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. Several 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 same way 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 thought that this field should continue to be displayed in US-ASCII and, to the extent possible, be displayed in the corresponding U-label.

[Editor note: The above may not be technically correct. Need to verify.]

[17] ICANN Registrar Accreditation Agreement, <http://www.icann.org/en/registrars/accreditation-application.htm>.

Telephone/Fax (RAA 3.3.1.7,8): The IRD-WG recommended that the UPU E.123 standard could be used to internationalize telephone and fax, specifically using the international notation (+31 42 123 4567). [18]

Email address (RAA 3.3.1.7,8): With email internationalization efforts ongoing, some IRD-WG members suggested that the email address field should be displayed according to latest the Internet Engineering Task Force (IETF) standard for international mail headers (draft-ietf-eai-rfc5335bis-11).

Dates (RAA 3.3.1.4,5): This includes creation date, expiration date, and update date of the domain. The IRD-WG members did not discuss the internationalization of this field.

[editor note: this needs to be resolved.]

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-ASCII7 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 discussed different opinions and chose option 2, since it gives client the ability to localize this field. Option 2 is also used in the new gTLD Draft Applicant Guide Book (DAG).[19]

[editor note: the above recommendation is gTLD centric. ccTLDs may not use EPP at all.]

Entity names and Address (RAA 3.3.1.6,7,8): This includes names and addresses of registrants, administrative contacts, and technical contacts. Recommendations concerning entity names will be discussed in detail in the next section.

5. Findings

5.1 Is it desirable to represent domain name registration data in non-US-ASCII?

Much of the original and currently accessible domain registration data are encoded in US-ASCII.

[18] International Telecommunications Union Recommendation E.123 <http://www.itu.int/rec/T-REC-E.123/en>.
[19] ICANN, "New TLD Program Application Guidebook," <http://www.icann.org/en/topics/new-gtlds/dag-en.htm>.

This legacy condition is convenient for WHOIS service users who are sufficiently familiar with languages that can be submitted and displayed in US-ASCII to be able to use ASCII script to submit domain name registration data and make and receive WHOIS queries using that script. Many millions of domain name registrants and Internet users do so today, even though their primary language is written using a different script. This condition is also convenient for registrants, registrars and registries and installed base of operational WHOIS services that display US-ASCII.

However, these data are less useful to the WHOIS service users who are only familiar with languages that require character set support other than US- ASCII. It is important to note that the latter (underserved) community is likely to continue growing and could outnumber the former in a matter of years.

Many registrants are monolingual, which is the expectation and motivation behind internationalized domain names[20]. Therefore, it is unreasonable to assume all of them know or can enter the registration data in languages other than their local language.

Thus, it is desirable to represent domain name registration data in non-US-ASCII.

However, this desirability should be balanced against other uses of the data. While a domain registrant 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 universally. (see http://forum.icann.org/lists/ird-wg-report/msg00004.html)

5.2 Is it feasible to introduce submission and display specifications to deal with IRD?

The IRD-WG believes that it is feasible to introduce submission and display specifications to deal with internationalized registration data, but there are several barriers that need to be overcome.

- Port 43 WHOIS protocol has not been internationalized: It imposes no formal encoding or composition requirements on data exchanged between a client and a server and hence no method of signaling encodings, and there may be different implementations from TLD to TLD on how this is done. This variation would create a non-uniform experience for users and interoperability problems. If ICANN continues to require Port 43 WHOIS, these issues need to be addressed urgently with the technical community. (NOTE: Solutions exist today to accommodate monolingual users in the web environment.) [SJM: what are they? Should they be described?]

[editor note: SJM's comments needs to be addressed.]

[20] See S. Hussain. <http://forum.icann.org/lists/ird-wg-report/msg00000.html>

- At a minimum, the current registration data needs to be tagged with language/script (RFC4646). In addition, the community would benefit from a standard registration data schema (for example in XML):

    - A formal data schema for registration data (for example in XML) would enable end user clients to better localize the data label and data.
    - A formal data schema for registration data with language tag information would allow better processing of the data.
    - Registrars may, in their policies, allow for multiple languages in the contact data (for example, they could allow for a Arabic registrant living in America to put his/her name in Arabic, but his/her address in English). If this is needed, the language-tag data needs to be at a data-element level.

- There are recognized standards for internationalizing many of the elements of domain registration data; to the extent possible, these standards should be followed. (See previous sections on these standards.)

- One key issue is internationalizing contact information. These data elements include Names (owner, admin, technical contact) and Addresses (owner, admin, technical contact).

    - The central issue here is to balance the needs and capabilities of the local registrant with the need of the (potential) global user of this data.

    - Several options have been discussed. These are:

        - Registrants provide domain contact data in "Must Be Present" script in addition to their local script.
        - Registrants provide data in any registrar-accepted script and registrars provide point of contact for transliteration or translation.
        - Registrants provide data in any registrar-accepted script and registrars provide transliteration tools to publish in "Must Be Present" script.
        - Registrants provide data in any registrar-accepted language and registrars provide translation tools to publish in "Must Be Present" script.
        - Registrants provide data in the script and language of their choice, but must specify language (locale(?)) of the data. There is no requirement for transliteration/translation of IRD (no "Must Be Present" script).

The WG examined these models and theirs impacts on registries, registrars, and end users. In the end, the WG could not reach consensus on which model to choose. This is the result of two reasons:

- Although there are standards in existence for transliteration and translation (see section 4); translation or transliteration process is highly subjective and there will be problems or

both consistency and accuracy:

- Transliteration or translation is not possible on the knowledge of the script alone. One would need to know the language.
- Transliteration or translation may vary significantly across languages using the same script.
- Two people may translate/transliterate differently even within a language.
- The same person may translate/transliterate differently at different times for the same language.
- As a concrete example, □□□□ is a commonly used name in the Arabic script based languages (270 million pages for □□□□ found on Google on 19th Feb. 2011). It is translated /transliterated in ASCII in the many ways (some listed below): Mohammed, Mohamed, Muhammed, Muhamed, Mohammad, Mohamad, Muhammad, Muhamad. So if □□□□ is the name of a monolingual registrant (a likely possibility), which spellings will Registrar A choose? Will Registrar B choose the same spelling? Also, how would a registrar determine which particular spellings to use for a particular registrant? How would the monolingual registrant ever verify such information even if presented such data by the registrar or by a third organization which does transliteration or translation?

- Secondly, if translation/transliteration is desired, there is a set of important policy questions that the composition of the working group could not tackle. These policy questions include: 1) Who should translate/transliterate, is it registrant, registrar or registry? 2) Who is in the best position to address this issue most effectively? 3) Who should bear the cost?

The WG offers the following notes based on past deliberations as well as public comments received:

- Registrants can be monolingual. As a result, responsibilities and roles for translation, transliteration, quality, and compliance need to be determined.

- We also note that ICANN's stated goals of maintaining stability and efficiency would benefit from a greater degree of uniformity in how gTLDs and ccTLDs handle internationalized registration data (and registration data in general). We encourage ICANN to consider how these goals could best be achieved through both technical standards adopted and policies developed.

[editor note: this last point does not seem to fit here.]

6. Recommendations

[editor note: These recommendations have not been discussed, WG members feel free to propose texts.]

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 should develop an issues report on translation and transliteration of contact information. The issues 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 report should consider policy questions raised in this document and should also determine whether to start a policy development process (PDP).

Recommendation 3: ICANN staff should work with the community to identify a directory service that meets the needs of internationalization and the service requirements enumerated in the report (WHOIS service requirements).

Appendix A: List of IRD-WG members

Appendix B: Different Models that IRD-WG considered for internationalizing contact information.

Model 1: Provide Directory Service Data in "Must Be Present" Script
Model 1 requires registrants to provide their directory service data in a "Must Be Present" script, for example, in US-ASCII7. The registrars have the option of asking registrants to provide their contact information in a local script. If registrants also provide information in their local script, then this information is displayed. Many IRD-WG members thought that that Model 1 was feasible because it has the least potential impact on registrars and registries. However, they also thought that it would provide the fewest benefits for internationalized registration data since local language display is optional. Figure 2 illustrates this model.

Figure 1: Model 1 for displaying contact information. In this model registrants provide data in US-ASCII7, and optionally in local script. The registrars display it in US-ASCII7.

Model 2: Provide Data in Registrar-Accepted Script and Point of Contact
In Model 2, registrants provide their registration data in a script that can be accepted by the registrar, and registrars provide a point of contact for transliteration and abuse issues on request. The registrars will also forward the same information to the registry. Many IRD-WG members also thought Model 2 was feasible. However, some IRD-WG members wondered whether this model would create inaccuracies. For example, in this model, registries may not verify the validity of the scripts they receive from registrants and may not take responsibility for the accuracy of the information. If the verification of the script is not performed, it is possible that an entry that combines Cyrillic, simplified Chinese, and Indic scripts could be created as a valid

WHOIS entry. In addition, some IRD-WG members were wary of any solution that relies upon registrar provision of a point of contact, whether to the public or to registrants. Figure 3 illustrates this model.

```
$ whois -h idnwhois.registrarX.ru    жук.рф
$ whois -h idnwhois.registrarX.ru  XN--F1AI0A.XN--P1AI

% Registrar X WHOIS server
% This query returned 1 object

domain:        жук.рф
domain-ace:    XN--F1AI0A.XN--P1AI
Registrar:     RU-CENTER LLC
Registrar POC: http://nic.ru
phone:         +7 800 234-5689
fax-no:        +7 800 234-5699
email:         info@nic.ru
contact:       Петр Иванв
organisation:  ОАО Циркуль
address:       ул.Ленина, офис 1, г.Ковров
address:       Владимирская обл. 601900
address:       Россия
phone:         +7 49232 48720
fax-no:        +7 49232 48722
e-mail:        cicle@cicle.ru
```

Figure 2: Model 2 to display contact information. Registrants in this model provide localized information and registrars provide a point of contact to respond to translation issues.

Model 3: Provide Data in Any Script Accepted by the Registrar; Registrar Provides Transliteration Tools to Publish in "Must be Present" Script

In Model 3, registrants provide their registration data in any script accepted by the registrar, and registrars provide tools for publishing the data in a "must be present" script. Many IRD-WG members raised concerns that Model 3 would incur added costs to registrars to produce transliterations. In addition, some IRD-WG members thought that transliteration would not be accurate enough to benefit law enforcement or intellectual property enforcement. Moreover, other members thought that Model 3 represents added value and that the focus on policy should be on baseline behavior, not on added value. Finally, some IRD-WG members were wary of any solution that relies upon registrar provision of transliteration services, whether to the public or to registrants. Figure 4 illustrates this model.

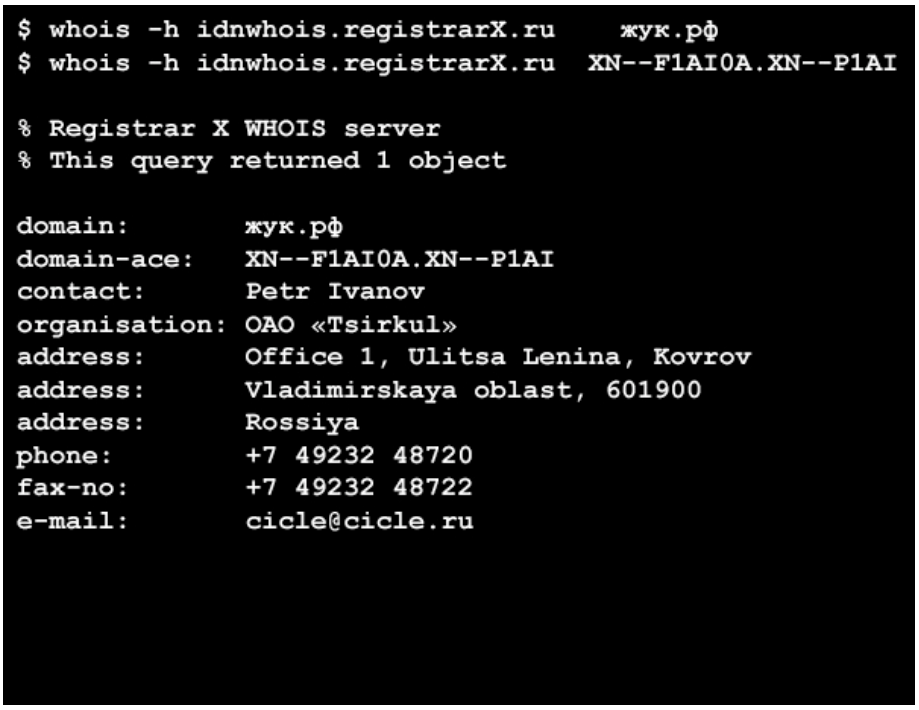

Figure 3: Model 3 to represent contact information. In this model, registrants provide information in local language, and registrars *transliterate* registrants' submission and display them in WHOIS.

Model 4: Provide Data in Any Script Accepted by the Registrar; Registrar Provides Translation Tools to Publish in "Must be Present" Script

In Model 4, registrants provide their registration data in any script accepted by the registrar, and registrars provide tools for translating and publishing the data in a "must be present" language. Many IRD-WG members raised concerns that Model 4 would incur added costs to registrars as they produce translations. In addition, some IRD-WG members thought that translation would not be accurate enough to benefit law enforcement or intellectual property enforcement. Moreover, other members thought that Model 4 represents added value and that the focus on policy should be on baseline behavior, not on added value. Finally, some IRD-WG members were wary of any solution that relies upon registrar provision of translation services, whether to the public or to registrants. Figure 5 illustrates this model.

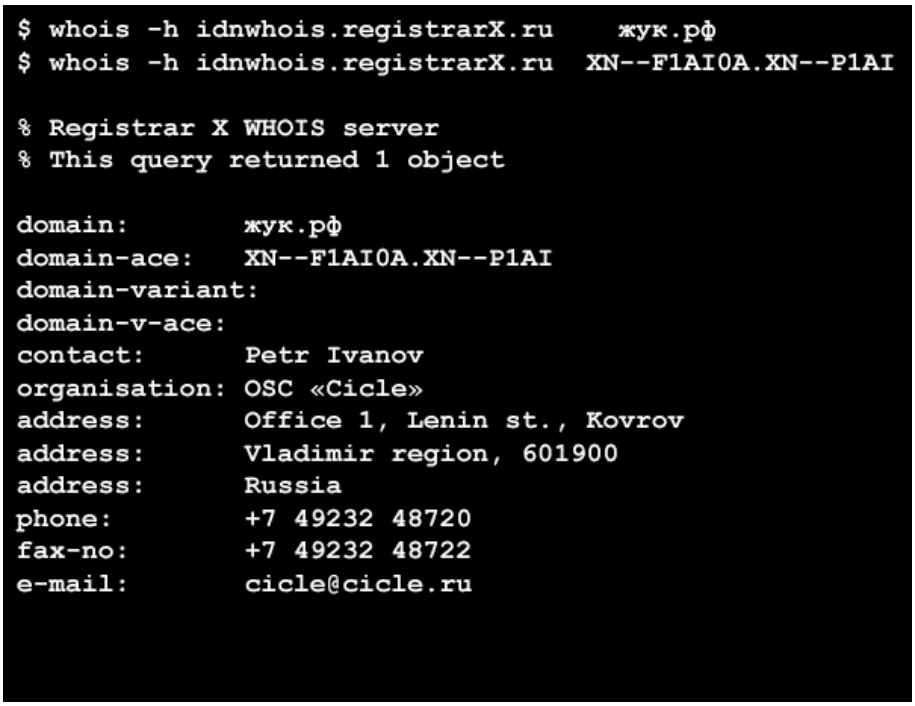

Figure 4: Model 4 to represent contact information. In this model, registrants provide information in their local language, and registrars *translate* registrants' submission and display them in WHOIS.

[editor note: Model 5 needs to be added.]