# Interim Report of the ICANN Internationalized Registration Data Working Group

**15 November 2010**

# Table of Contents

# 1. Executive Summary

On 26 June 2009 the Board of Directors of the Internet Corporate 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) 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 further directed the IRD-WG to solicit input from interested constituencies including country code top level domain (ccTLD) operators and the Country Code Names Supporting Organization (ccNSO) during its discussions to ensure broad community input.[1] Subsequently, the SSAC and the GNSO formed the IRD-WG. This interim report of the IRD-WG summarizes the discussions of the IRD-WG to date, provides preliminary recommendations, and provides questions for community discussion.

# 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]

Access to domain name registration information (often called Whois[3] data) is provided by several applications. Accommodating the submission and display of internationalised 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

[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.

[3] In this report, "Whois" in lower case with initial capital letter refers to Whois services in general. "WHOIS" in all capital letters is used to reference the WHOIS protocol (RFC 3912 and earlier versions).

> [Daigle 2004][4] 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."[5]

The SSAC called attention to these issues in SAC037, *Display and usage of Internationalized Registration Data, Support for Characters from Local Languages or Scripts*.[6] In the report the SSAC recommended that the ICANN Board of Directors should 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.

This interim report of the IRD-WG summarizes the discussions of the IRD-WG to date, identifies a list of remaining issues, including preliminary recommendations and questions for community discussion.

# 3. Terminology, Background, and IRD-WG Objectives and Goals

## 3.1 Background

ICANN requires that a registrant provide certain information when registering a domain name, and that registrars or registries make this registration information available for public scrutiny. The Registrar Accreditation Agreement (RAA 3.3.1) specifies the following data elements that must be provided by registrars in response to a query:

3.3.1.1 The Registered Name;

3.3.1.2 The names of the primary name server and secondary name server(s) for the Registered Name;

3.3.1.3 The identity of the Registrar (which may be provided through Registrar's website);

3.3.1.4 The original creation date of the registration;

3.3.1.5 The expiration date of the registration;

3.3.1.6 The name and postal address of the Registered Name Holder;

---

[4] 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>.

[5] 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>.

[6] 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>.

3.3.1.7 The name, postal address, e-mail address, voice telephone number, and (where available) fax number of the technical contact for the Registered Name; and

3.3.1.8 The name, postal address, e-mail address, voice telephone number, and (where available) fax number of the administrative contact for the Registered Name.

Some registry agreements include a more extensive list of data elements that must be displayed. [7]

Various protocols have been developed to support the query and display of domain name registration data. Among them, ***the original WHOIS protocol*** (RFC 3912) is most widely used.[8] The protocol 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 American Standard Code for Information Interchange *(*ASCII) line feed (LF) and carriage return (CR) character sequence. Other protocols have been developed or used to transport domain registration data, including IRIS[9] and Whois-RWS,[10] based on Representational State Transfer (RESTful) .[11]

## 3.2 Terminology

The term "Whois" has several meanings and connotations. There are three different uses of the term at ICANN.

1. ***The WHOIS protocol, RFC 3912.*** In this report, we use all capital letters when referring to the WHOIS protocol.
2. ***The Whois services.*** These are a set of applications that provide access to domain name registration information via the WHOIS protocol and web based interfaces. Most non-technical users access the Whois services via the web based interfaces. In this report, we capitalize the word "whois" when referring to the Whois services.
3. ***The data.*** This refers to the domain name registration information that is collected at the time of registration and periodically updated. This information is made available to users and automation services via Whois services according to the Registrar Accreditation

[7] See for example <http://www.icann.org/en/tlds/agreements/biz/appendix-05-08dec06.htm.
[8] 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>.
[9] A. Newton and M. Sanz, "RFC 3981: IRIS: The Internet Registry Information Service (IRIS) Core Protocol," Network Working Group, Internet Enginerring Task Force, The Internet Society, January 2005, <http://www.ietf.org/rfc/rfc3981.txt> and A. Newton and M. Sanz, "RFC 3982: IRIS: A Domain Registry (dreg) Type for the Internet Registry Information Service (IRIS)," Network Working Group, Internet Enginerring Task Force, The Internet Society, January 2005, <http://www.ietf.org/rfc/rfc3982.txt>.
[10] American Registry for Internet Numbers (ARIN), "Whois-RWS API Documentation," <https://www.arin.net/resources/whoisrws/whois_api.html>.
[11] Wikipedia, "Representational State Transfer," <http://en.wikipedia.org/wiki/Representational_State_Transfer>.

Agreement (RAA). In this report, we refer primarily to ***registration data***, and in particular, to ***internationalized registration data (IRD).***[12]

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 consistent with ICANN's IDN glossary.[13]

***Internationalised 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 "a-z." 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.

***Internationalised 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 hyphen (-). By registration data elements, we mean data such as contact information, host names, sponsoring registrar, and domain name status.

***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 (e.g. □□□□□) to produce a corresponding A-label (e.g. xn--11b5bs1di). The transformation is symmetric, i.e., 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.

***Variant characters:*** Variant characters (as defined in RFC 3743) occur where a single conceptual character has two or more graphic representations, which may or may not be visually similar.[14]

***IDN variant:*** is an IDN that includes one or more variant characters in the label.

---

[12] It is important to note that for the purpose of our discussion, we consider registration data separately from the WHOIS protocol and Whois services.
[13] ICANN, "IDNs Glossary," Retrieved August 10, 2010, <http://www.icann.org/en/topics/idn/idn-glossary.htm>.
[14]K. Konishi, H. Qian, and K. Huang, "Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean," Network Working Group, Internet Engineering Task Force, Internet Society, April 2004, <http://www.ietf.org/rfc/rfc3743.txt>.

## 3.3 IRD-WG Objectives and Goals

***The first objective of the IRD-WG is to identify how to internationalize domain registration data.*** Currently, IDN guidelines define how IDNs will be composed and displayed.[15,16] Application and web developers can apply these standards for submission and display of internationalised domain names. However, no standard or guidelines define how internationalized domain registration data are to be composed and displayed. The data include registrant contact information, host names, sponsoring registrar, and domain name status, hereinafter referred to as internationalized registration data (IRD).

***The second objective of the IRD-WG is how to specify how to internationalize the WHOIS protocol.*** Today, certain WHOIS protocol implementations are able to transport IRD, but do so in a non-standard, non-uniform manner. The Internet Standard for the protocol does not specify a character set. It also does not specify a mechanism for a client to indicate, propose, or request a character set to a server, or for the server to indicate character set support. The inability to predict or a express character set (encoding) has an adverse effect on the interoperability (and, therefore, usefulness) of the WHOIS protocol.

***The goals of the IRD-WG are to:***

- Study the feasibility and suitability of introducing submission and display specifications to deal with the internationalization of Registration Data; and
- Engage participation from all ICANN Supporting Organizations and Advisory Committees as well as Country Code top level domain (ccTLD) operators, to ensure broad community input.

***IRD-WG Membership:*** Edmon Chung (GNSO) and Jeremy Hitchcock (SSAC) co-chair the IRD-WG. The international representation in the IRD-WG includes 17 participants, 5 staff support, 5 countries (China, Tunisia, New Zealand, Russia, and the USA), 3 ccTLDs (.cn, .nz, .ru)\ and 3 ICANN Supporting Organizations and Advisory Committees (ALAC, GNSO, SSAC).[17]

---

[15] P. Fältström, P. Hoffman, P., and A. Costello, "RFC 3490: Internationalizing Domain Names in Applications," <http://www.ietf.org/rfc/rfc3490.txt>; A. Costello, RFC 3492: Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications," <http://www.ietf.org/rfc/rfc3492.txt>; and P. Hoffman, and M. Blanchet, "RFC 3491: Nameprep: A Stringprep Profile for Internationalized Domain Names," <http://www.ietf.org/rfc/rfc3491.txt>, Network Working Group, Internet Engineering Task Force, Internet Society, March 2003

[16] P. Hoffman, and M. Blanchet, M. "RFC 3454: Preparation of Internationalized Strings," Network Working Group, Internet Engineering Task Force, Internet Society, December 2002, <http://www.ietf.org/rfc/rfc3454.txt>

[17] 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>.

# 4. Summary of IRD-WG Discussions

The IRD-WG discussed a number of issues relating to internationalization of registration data. Of these issues, two were considered overarching areas of interest:

1. The deficiencies identified by the IRD-WG of the WHOIS protocol; and
2. The query and display of variants.

The IRD-WG members also discussed the question, "What capabilities are needed for directory services in the internationalized domain name and registration environment?" and considered various aspects of the question, "How should Whois services present registration data in different scripts?" Deliberation of these questions resulted in the identification and description of four possible models and their respective impacts on users, registrants, registrars and registries.

## 4.1 The Deficiencies of the WHOIS Protocol and Whois Services

Members of the Internet and ICANN communities express concern that the current WHOIS protocol does not meet the community's current and future needs. These are noted in recent reports from the SSAC,[18] in staff reports to other ICANN supporting organizations and advisory committees[19] and by external sources.[20] With respect to internationalization, the deficiency is described in the protocol specification itself in RFC 3912:

> "The WHOIS protocol has not been internationalised. The WHOIS 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."[21]

Whois services serve many kinds of users. However, as we have noted, the increased desire and need for Internet applications to accommodate users who use scripts that are not based solely on the US-ASCII character set exposes the following challenges:

- ❑ Text requests and content returned by Whois services are historically encoded using US-ASCII7. While the WHOIS protocol does not specify US-ASCII7 as the exclusive character set for text requests and text content encoding, and thus gives

[18] Security and Stability Advisory Committee, "SAC003: WHOIS Recommendation of the Security and Stability Advisory Committee,"<http://www.icann.org/en/committees/security/sac003.pdf>, "SAC027, Comment to GNSO regarding WHOIS studies," <http://www.icann.org/en/committees/security/sac027.pdf>, and "SAC033, Domain Name Registration Information and Directory Services," <http://www.icann.org/en/committees/security/sac033.pdf>

[19] ICANN Generic Names Supporting Organization (GNSO) "Inventory of Whois Service Requirement Final Report," 31 May 2010, <http://gnso.icann.org/issues/whois/whois-service-requirements-draft-final-report-31may10-en.pdf>.

[20] A. Newton, "Replacing the WHOIS Protocol: IRIS and the IETF's CRISP Working Group," Internet Computing, IEEE Volume: 10 Issue: 4 July-Aug. 2006 Page(s): 79-84.

[21] 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>.

latitude with respect to protocol encoding, the protocol specification leaves the method of signalling/selecting character sets as a local implementation matter. The current situation is that no standards or conventions exist for all WHOIS protocol implementations to signal support of character sets other than US-ASCII. Whois services are supported by a large and diverse set of providers for an even larger and more diverse set of users who increasingly want to access Whois using a familiar script or language; thus, the lack of a signalling convention is problematic.

- Much of the original and currently accessible domain registration data are encoded in US-ASCII7. This legacy condition is convenient for Whois service users that are familiar with languages that can be submitted and displayed in US-ASCII7. It is also convenient for registrants, registrars and registries and installed base of operational Whois services that display US-ASCII7. However, these data are less useful to the Whois service users that are only familiar with languages that require character set support other than US- ASCII7. It is important to note that it is very likely that the latter (underserved) community will continue to grow and could outnumber the former in a matter of years.
- Much of the automation developed to parse and analyze domain registration data assumes that the registration data element labels and the data proper are encoded in US-ASCII7. Increasingly, applications that make these assumptions will not process all registration record data in the manner intended. (We acknowledge that this is one of several issues related to the non-uniformity of registration data across registries today, but it will become an increasingly troublesome issue over time).
- The ACE method for encoding internationalized domain names to provide backwards compatibility in the DNS protocol cannot be generalized to accommodate the encoding of all registration record data. The issue for Whois services is not simply one of preserving backwards compatibility but a more general matter of defining an extensible framework for character set selection and transport between a client and server application.
- The introduction of IDNs creates the need to consider certain data elements beyond the current set identified in the ICANN RAA, e.g., variants. How to best support extensible data is an important consideration for the IRD-WG.
- The most beneficial resolution of IRD is one that will be widely adopted by both gTLD registries and ccTLD registries and thus the development of conventions or policy requires participation and cooperation from a very broad stakeholder community.

## 4.2 Query and Display of Variants in Internationalized Registration Data

Variant characters occur where a single character has two or more representations, which may or may not look visually similar. For example, in CJK (Chinese, Japanese, Korean), the term “international” can have several different code points. In Chinese it can be written in simplified

Chinese as 国际, or 國際 in traditional Chinese. In Japanese it can be written as 国際, but 囡際 is also acceptable.[22]

In some languages such as Chinese, simplified Chinese (SC) and traditional Chinese (TC) are treated with equivalence. As another example, the variants for IDN label **清华大学** (Tsinghua University) will include: **清华大學、清華大学、清華大學、□华大学、□华大學、□華大学、□華大學**.[23]

The IRD-WG members discussed the issue of how to query and display variants extensively. They provide the following observations:

- There is no uniform definition of variant. Different organizations and different countries define it differently. However, in general, variants can be categorized as *activated* variants and *reserved* variants. Activated variants are variants of a domain name that are put in the corresponding DNS zone file, thus resolvable through normal DNS lookups. Reserved variants are variants reserved for a specific domain name and cannot be registered, but are otherwise not in the DNS zone file.
- IRD-WG members noted that it is outside the scope of the IRD-WG to define variants or discuss how different languages handle variants. Rather, the IRD-WG use the categories as they are generallly defined (activated vs. reserved).
- The IRD-WG members agree that a Whois service query of an activated variant should return the domain of which it is a variant in its response, as well as an indication that the label queried is a variant of the original domain. The IRD-WG members agree that this should be consistent across Whois services.
- The IRD-WG members also agree that defining a Whois service query of a reserved variant returns is a matter of local policy. The IRD-WG has identified two options: A query of a reserved variant for XYZ domain should return a message saying that this variant is a reserved variant of XYZ domain or (alternatively) a query of a reserved variant should return the same information as the query for an activated variant. The WG further agreed that having the Whois service response provide a link to the registrar/registries' variant policy would be helpful.

## 4.3 What Capabilities Are Needed for Directory Services in the IDN Environment?

The IRD-WG discussed the question "What is an appropriate (satisfactory) user experience when a user submits an IDN as a query argument to a Whois service?"

[22] J., Yao, "RFC 3743 and IDN TLD tests," CNNIC. Retrieved August 30, 2010 from <http://losangeles2007.icann.org/files/losangeles/ChinaonIDNs.pdf>.
[23] Ibid.

The IRD-WG members agree that there is value in supporting the ability to submit either a U-label (Unicode form of an IDN label) or an A-label (ASCII form of an IDN label) as a query argument to a Whois service. Users may most often prefer a U-label (e.g. 测试.test) since this is more visually recognizable and familiar than A-label strings (e.g. or XN--0ZWM56D.test), but users of programmatic interfaces may want to submit and display A-labels or may not be able to input a U-label with their keyboard configuration.

For illustration, below is a screenshot of a WHOIS service that met the above requirements for a fictitious IDN domain 测试.test.

```
$ whois -h new.whois.registrarX.com 测试.test
$ whois -h new.whois.registrarX.com XN--0ZWM56D.test

% Registrar X WHOIS server
% This query returned 1 object

domain:       测试.test
domain-ace:   XN--0ZWM56D.test
domain-variant: 測試.test
domain-v-ace: XN--G6W251D.test

organisation: Internet Assigned Numbers Authority
address:      4676 Admiralty Way
address:      Suite 330
address:      Marina del Rey California 90292
address:      United States

...
```

**Figure 1: Sample WHOIS output for domain 测试.test that conforms to the recommended Whois service requirements. In this illustration, a user can submit either the query 测试.test (Simplified Chinese U-label) or XN--0ZWM56D.test (corresponding A-label) and get the same result back. The Whois displays both A-label and U-label representation of the domain as well as its traditional Chinese variant 測試.test (XN--G6W251D.test).**

## 4.4 How to Accommodate Users Who Want To Submit and Have Registration Data Displayed in Local Scripts

The IRD-WG members agreed that various elements of registration data could be separately internationalized. (See Section 3.1 Background above.)

***Domain names (RAA 3.3.1.1):*** Whois services should return both A-label and U-label representation for the given IDN domains queried.

***Name server names (RAA 3.3.1.2):*** Currently all name servers are in US-ASCII. However, with IDNs, it is possible that some registrants will compose name server names using IDN labels. Several alternatives exist:

1. Always display the name server name in US-ASCII 7 using the A-label. A supporting argument for this choice is that name server name information is generally only of technical interest and should be displayed in same way as it is in the DNS; and
2. Display name server 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-ASCII7, and to the extent possible be displayed in the corresponding U-label.

***Sponsoring Registrar (RAA 3.3.1.3):*** The IRD-WG members thought that this is an example of data that should always be available 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."[24]

***Telephone/Fax (RAA 3.3.1.7,8):*** Some IRD-WG members suggested that the UPU E.123 standard could be used to internationalize telephone and fax, specifically using the international notation (+31 42 123 4567). [25]

***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 the Internet Engineering Task Force (IETF) standard for international mail headers (RFC 5335 now, and according to successor RFCs as they are completed).

***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.

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

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; or
3. Identify a more easily understood representation;

[24] ICANN Registrar Accreditation Agreement, <http://www.icann.org/en/registrars/accreditation-application.htm>.
[25] International Telecommunications Union Recommendation E.123 <http://www.itu.int/rec/T-REC-E.123/en>.

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).[26]

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

## 4.5 Models for Internationalizing Registration Contact Data

One of the key questions the IRD-WG members discussed is whether directory services should support multiple representations of the same registration data in different languages or scripts. In particular, the IRD-WG members discussed whether it is desirable to adopt a "must be present" representation of data, in conjunction with local script support for the convenience of local users. The IRD-WG has identified four models for internationalizing registration data such as contact information that includes registrant name, administrative contact, technical contact, and postal addresses. The IRD-WG members determined that they need guidance from the ICANN and international community on the feasibility of the four models in order to gain the additional information they will need to develop specific recommendations. Questions for the community to consider are described below in Section 6 Questions for Community Discussions.

### 4.5.1 Model 1: Provide Directory Service Data in "Must Be Present" Script

Model 1 would require registrants to provide their directory service data in a "must be present" script, for example, in US-ASCII7. Optionally, the registrars could also ask registrants to provide their contact information in a local script. If registrants also provide information in their local script, then this information should be 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.

[26] ICANN, "New TLD Program Application Guidebook," <http://www.icann.org/en/topics/new-gtlds/dag-en.htm>.

```
$ 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
domain-variant:
domain-v-ace:
contact:      Petr Ivanov (Петр Иванов)
organisation: OSC «Cicle»
address:      Office 1, Lenin st., Kovrov
address:      Vladimir region, 601900
address:      Russia
phone:        +7 49232 48720
fax-no:       +7 49232 48722
e-mail:       cicle@cicle.ru
```

**Figure 2: 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.**

### 4.5.2 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 script 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 3: 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.**

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

In Model 3, registrants would provide their registration data in any script accepted by the registrar, and registrars would provide tools to assist the registrant so it can be published it in a "must be present" script. Many IRD-WG members raised concerns with respect to Model 3 because of the added cost 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 behaviour, 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.

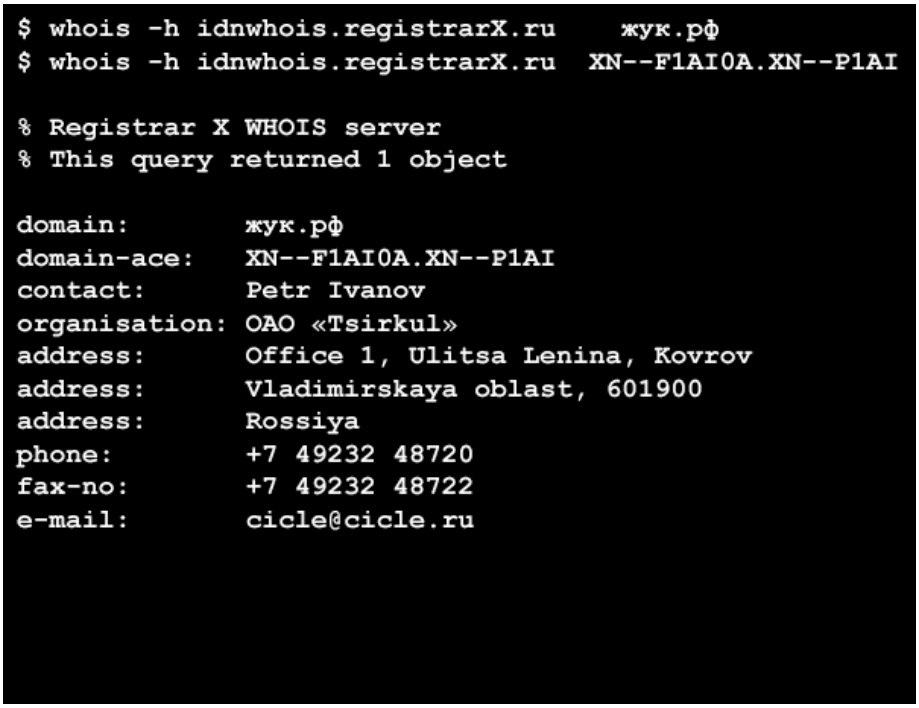


**Figure 4: Model 3 to represent contact information. In this model, registrants provide information in local language, and registrars *transliterate* registrants' submission and display them in Whois.**

### 4.5.4 Model 4: Provide Data in Any Script Accepted by the Registrar and 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 registrar provides tools to assist the registrant translating and publishing it in a "must be present" language. Many IRD-WG members raised concerns with respect to Model 4 because of the added cost to registrars to 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 behaviour, 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.

```
$ 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
domain-variant:
domain-v-ace:
contact:      Petr Ivanov
organisation: OSC «Cicle»
address:      Office 1, Lenin st., Kovrov
address:      Vladimir region, 601900
address:      Russia
phone:        +7 49232 48720
fax-no:       +7 49232 48722
e-mail:       cicle@cicle.ru
```

**Figure 5: Model 4 to represent contact information. In this model, registrants provide information in local language, and registrars *translate* registrants' submission and display them in Whois.**

### 4.5.5 Impact of the Models

The IRD-WG members discussed the impact of the four models on registrars, registries, users of Whois, and the Whois service itself. The following tables summarize the discussions.

**Table 1: Impact of the Models on Existing Registrars**

| **Impact on Existing Registrars** | | | |
|---|---|---|---|
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| Use US ASCII-7 as the default script. If they accept additional language inputs, there is the cost to support the additional script because registrars need to store the data and update the registration process. | Provide a point of contact to address transliteration and abuse issues for each script in which they accept registrations. This increases the registrars' cost. It is also important to set some service guarantees for this point of contact, otherwise a transliteration or abuse request could be delayed. | 1) Increases the cost for registrars because they must provide the transliteration service; 2) creates uncertainty: who is responsible for the accuracy of the transliteration? | 1) Increases the cost for registrars because they most provide the translation service; 2) creates uncertainty: who is responsible for the accuracy of the translation? |

## Table 2: Impact of the Models on New IDN-Based Registrars

| **Impact on New IDN-Based Registrars** | | | |
|---|---|---|---|
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| Registrars will be required to put multi-script pages on the front end; and 2) if they are to require local scripts in addition to US-ASCII they will ask registrants to provide that information. | Provide a point of contact to address transliteration and abuse issues for each script in which they accept registrations. This increases the registrars' cost. It is also important to set some service guarantees for this point of contact, otherwise a transliteration or abuse request could be delayed. | 1) Increases the cost for registrars because they must provide the transliteration service; 2) creates uncertainty: who is responsible for the accuracy of the transliteration? | 1) Increases the cost for registrars because they most provide the translation service; 2) creates uncertainty: who is responsible for the accuracy of the translation? |

## Table 3: Impact of the Models on Thin Registries

| **Impact on Thin Registries** | | | |
|---|---|---|---|
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| No impact, as only US-ASCII7 is submitted. | No impact. | No impact. | No impact. |

## Table 4: Impact of the Models on Thick Registries

| **Impact on Thick Registries** | | | |
|---|---|---|---|
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| No impact. | The registries will not be able to interpret the registrant information unless they have a service that can translate the script. This will prevent them from administering a domain or from extracting detailed statistical information. It may also hinder them when looking for similar data in different registrations for such purposes as abuse detection. Under the current model of gTLDs and most ccTLDs there is no official role for the take these actions, but if such a role were to be developed then this would prevent it. | No impact. | No impact. |

## Table 5: Impact of the Models on Registrants

| **Impact on Registrants** | | | |
|---|---|---|---|
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| Some barrier of entry for registrants, as they have to know or find someone to transliterate the script for them. If the registration information is not checked, this could lead to inaccuracies in Whois. | Least barrier of entry to registrants. | Least barrier of entry to registrants. | Least barrier of entry to registrants. |

## Table 6: Impact of the Models on Whois Service Users

| **Impact on Whois Service Users** | | | |
|---|---|---|---|
| **Only Users of Internationalized Registration Data** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| May enhance usability of Whois because the data is in local script. | May enhance usability of Whois because the data is in local script. | May enhance usability of Whois because the data is in local script. | May enhance usability of Whois because the data is in local script. |
| **ASCII-Only Capable User** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| Unchanged, because there will be a "must be present" script. | Poses significant challenges because Whois data would be displayed in many scripts that the local user would not understand. | Unchanged, because there will be a "must be present" script. | Unchanged, because there will be a "must be present" script. |
| **Legitimate Automation User** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| There would be little impact. | There would be little impact as the encoding would likely be UTF-8. | There would be little impact. | There would be little impact. |

## Table 7: Technical Impact on Applications: Port 43 Whois Clients

| **Technical Impact on Applications** | | | |
|---|---|---|---|
| **Port 43 Whois Clients** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| The WHOIS protocol has no mechanism for indicating the character set in use. In the absence of protocol solution, some Whois servers that support IRD require flags to specify output encodings (e.g. .DE, .JP WHOIS servers). It may be possible to specify input encodings, but this is unclear. Also the terminal that runs the Whois client must have the same encoding as the Whois server output to display properly (for example ISO-2022-JP). | The impact is unclear. | The impact is unclear. | The impact is unclear. |
| **Web Whois Clients** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| This depends on both the operating system and the browser. For characters to be displayed properly, the browser needs to support Unicode. Moreover, an appropriate Unicode font must be available to the browser. Often, Unicode fonts do not display all the Unicode characters. Some platforms provide partial support for Unicode. | The impact is unclear. | The impact is unclear. | The impact is unclear. |

**Table 8: Technical Impact on Applications: Other Whois Clients**

| **Technical Impact on Applications** | | | |
|---|---|---|---|
| **Other Whois Clients** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| The impact is unclear. | The impact is unclear. | The impact is unclear. | The impact is unclear. |
| **Impact on the Whois Service** | | | |
| **Model 1:US-ASCII Mandatory; Local Script Optional** | **Model 2: Accept and Display Any Script; Provide Point of Contact** | **Model 3: Accept any Script; Provide Transliteration** | **Model 4: Accept any Script; Provide Translation** |
| The impact is unclear. | The impact is unclear. | No impact. | No impact. |

### 4.5.6 Harmonizing Registration Data Labels Across Whois Services

The previous discussion focused on IRD. Separately, a question arises about whether or not to internationalize (or localize) the labels for these data elements in Whois services. Specifically, should the labels always be in US-ASCII, or should it be completely localized?

Regarding this point, some IRD-WG members identified several alternatives:

- The labels should be in US-ASCII by default;
- ICANN should harmonize the labels used in different registries and registrars; and
- Translation of labels to other languages may be accomplished by string replacement tables maintained by ICANN or IANA.

# 5. Preliminary Recommendations for Community Discussion

The IRD-WG offers the following preliminary recommendations for community consideration. The IRD-WG welcomes the community's input on this Interim Report, and will use the input as well as the continued deliberations in the IRD-WG to reach a set of recommendations.

***Preliminary Recommendation (1):*** The IRD-WG discussed a preliminary recommendation for a Whois service in the IDN environment:

1. WHOIS clients (both port 43 and web) must be able to accept a user query of domain name in either U-label or A-label format;
2. WHOIS clients must be able display result of queries in both U- and A-label for the domain names; and

3. Whois responses should include variants of an IDN label in the response as well.

***Preliminary Recommendation (2):*** The IRD-WG discussed the idea that the domain registration data elements should be considered separately, with specific recommendations for how each data element should be internationalized. The IRD-WG offers preliminary recommendations for the following data elements:

1. Whois services should return both A-label and U-label representation for the given IDN domains queried;
2. Whois services should return both A-label and U-label representations for nameserver names (to the extent that such information is available);
3. Whois services should always make sponsoring registrar information available in US-ASCII7; and
4. Whois services should always return the exact EPP[27] status code for Registration Status.

# 6. Questions for Community Discussion

The IRD-WG calls attention to and seeks comment the following questions regarding internationalized registration data:

1. Which of the four models described in Section 4 for internationalizing registration contact data is most appropriate, if any? Are there other models the IRD-WG should consider?
2. Which of the preliminary recommendations in Section 5, if any, are feasible. Are there related recommendations the IRD-WG should consider?

# 7. Summary

This interim report of the IRD-WG summarizes the discussions of the IRD-WG to date, identifies a list of remaining issues, including preliminary recommendations and questions for community discussion. The Interim Report provides the community with an opportunity to understand ongoing discussions relating to the internationalization of registration data and to provide valuable input to these discussions.

[27] Wikipedia, "Extensible Provisioning Protocol (overview)," <http://en.wikipedia.org/wiki/Extensible_Provisioning_Protocol>.