

# Draft Interim Report of Internationalized Registration Data Working Group

## STATUS OF THIS DOCUMENT

This is an interim report for the Internationalised Registration Data Working Group.

## SUMMARY

## TABLE OF CONTENTS

<b>1</b>	<b>EXECUTIVE SUMMARY</b>	<b>3</b>
<b>2</b>	<b>INTRODUCTION</b>	<b>3</b>
<b>3</b>	<b>BACKGROUND AND TERMINOLOGY</b>	<b>4</b>
<b>4</b>	<b>ISSUES</b>	<b>7</b>
<b><u>5</u></b>	<b>NEXT STEPS</b>	<b><u>20</u></b>

# 1 Executive Summary

To address the issues of internationalised registration data in Whois, The ICANN security and stability committee (SSAC) and the Generic Names Supporting Organization (GNSO) formed a joint working group to study the feasibility and suitability of introducing submissions and display specifications to deal with the internationalised registration data. This report is the interim report of the Working Group, it summarizes the discussions of the Working Group to date, compiles a set of recommendations from discussions, and identifies a list of remaining issues.

## 2 Introduction

With the increasing use of the Internet in all geographic regions and by diverse linguistic groups of the world, the demand for a multilingual Internet has become more intense. Many Internet applications are now able to accept as input and display characters from a broad range of languages and scripts. 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.<sup>1</sup>

As the Internet is becoming more internationalised, questions arise about the domain registration database (Whois<sup>2</sup> database) and the protocol (WHOIS protocol) to support this. The following statement from the Internet Architecture Board (IAB), found in RFC 4690, summarizes the issues:

“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] 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.” [RFC 4690]

---

<sup>1</sup> 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.

<sup>2</sup> “Whois” is used in reference to the service in general and “WHOIS” (in caps) is used when referring to the WHOIS protocol (RFC 3912) and older protocol.

Recognizing the problem, the ICANN Security and Stability Advisory Committee (SSAC) called attention to issue in SSAC 037 [ICANN SSAC 2009a] recommending that the ICANN Board of Directors for a working group to study the feasibility and suitability of introducing submissions and display specifications to deal with the internationalization of registration data. At the request of the ICANN Board of Directors, the Generic Names Supporting Organization (GNSO) and the SSAC created an Internationalized Registration Data Working Group (IRD WG) to study this issue.

This report is the interim report of the IRD Working Group. It summarizes the discussions of the Working Group to date, compiles a set of recommendations from discussions, and identifies a list of remaining issues. It is organized as follows: In section 3, we present background information on internationalized registration data. In section 4, we identify a list of issues that the WG considered, the findings that resulted from the discussions and the recommendations for each issue.

## 3 Background and Terminology

### 3.1 Whois Service

When people refer to Whois, they may mean several different things. There are three different uses of the word "WHOIS" at ICANN.

- (1) WHOIS - refers to the WHOIS protocol - RFC 3912
- (2) WHOIS -refers to the WHOIS "service" - which provides information via both the WHOIS protocol and web based interfaces. Most non-technical users access WHOIS via the web based interfaces.
- (3) WHOIS - refers to the data that is collected at registration and made available via the WHOIS service per Registrar Accreditation Agreement (RAA)

**For the purpose of our discussion, we consider the data and protocol separately:**

**Whois data** refers to the registration data that registrants provide and registrars or registries disclose. 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 nameserver and secondary nameserver(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;
- 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.

Various protocols have been developed to support the query and display of Whois data. Among them, **the original WHOIS protocol** (RFC 3912) is most widely used. The protocol describes exchanges of queries and messages between a client and a server over a specific port (43). It is very simple as the only constraint imposed on queries and messages is that they are terminated using an ASCII line feed (LF) and carriage return (CR) character sequence. In addition to the existing WHOIS protocol, other protocols have been developed or used, among them IRIS [Newton 2005a, 2005b] and RESTful Whois (RWS).

Currently, Internationalized Domain Name (IDN) guidelines define how internationalised domain names will be composed and displayed [Costello, 2003, Falstrom 2003, Hoffman 2002, 2003]. Whois application and web developers can apply these standards for submission and display of internationalised domain names. However, no standard of guidelines define how Whois data are composed and displayed. The data include registrant contact information, host names, sponsoring registrar, and domain name status, hereinafter termed as *internationalized registration data* (IRD). **Thus the first objective is to identify how to internationalize the Whois data.**

Furthermore, the WHOIS protocol [Daigle 2004] has not been internationalized in a standard or uniform manner. The Internet Standard for the protocol does not specify a character set. It does not specify a mechanism for a client to indicate, propose, or request a character set to use either. This inability to predict or express text encoding has adversely impacted the interoperability (and, therefore, usefulness) of the WHOIS protocol. **Thus the second objective is how to specify how to internationalize the WHOIS protocol.**

## 3.2 IRD WG Goals

Recognizing the problem, SSAC studied this issue in SSAC037 [ICANN SSAC 2009a]. The SSAC report examines how the use of characters from local scripts currently affects the Internet user experience with respect to domain name registration data submission, usage and display. At SSAC's recommendation, ICANN's Board of the directors tasked GNSO and SSAC to form the Internationalized Registration Data

Working Group (IRD WG) to study the feasibility and suitability of introducing display specifications or standards to deal with the internationalization of Registration Data.

The goals of the IRD WG are:

- 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.

The WG is chaired by Edmon Chung (GNSO) and Jeremy Hitchcock (SSAC). The international representation in the IRD-WG includes 17 participants, 5 staff support, 5 countries (China, Morocco, New Zealand, Russia, USA), 3 ccTLDs (.cn, .nz, .ru) and 3 SOs/ACs (ALAC, GNSO, SSAC).

### 3.3 Terminology

In an attempt to ensure that discussions regarding internationalised 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 [ICANN 2007].

**IDN (Internationalised domain name):** IDNs are domain names that include characters used in the local representation of languages 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.

**IRD (Internationalised Registration Data):** 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" (abbreviated ACE) is applied to a U-label (e.g. 中国) 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.

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

## 4 Challenges

[comment: I do not understand the following paragraph] 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 languages that are not based solely on the US-ASCII character set exposes several challenges:

- ❑ Text requests and content returned by Whois services are historically encoded using US-ASCII7; however the Whois protocol does not specify US-ASCII7 as the exclusive character set for text requests and text content encoding. While the protocol thus gives latitude with respect to protocol encoding, it leaves the method of signalling/selecting character set as a local implementation matter. Given that 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, the lack of a signalling convention is problematic.
- ❑ Much of the original and current registration record data accessible via Whois services is encoded in US-ASCII7. This legacy condition remains convenient for the Whois user community that is familiar with languages that can be submitted and displayed in US-ASCII7. It is also convenient for registrants, registrars and registries with experience in operating within the current system. Characterizing this legacy condition as satisfying the criteria for a vehicular language is, however, incorrect. These data have been less useful to the Whois user community that is restricted to languages that require character set support other than US-ASCII7; more importantly, it is very likely that the latter community will grow and could outnumber the former in a matter of years.
- ❑ Much of the automation developed to parse and analyze registration record data assumes that the 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, 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 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 internationalizing registration data 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.

## 4.1 What capabilities are needed for Whois service in the IDN environment?

The first question the IRD WG discussed is what kind of user experience should a user have when he or she query Whois with IDNs.

IRD WG members agree that there is value in supporting the the ability to "use" either a U-label (Unicode form of an IDN label) or an A-label (ASCII form of an IDN label) in WHOIS queries. 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—0ZW56D.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.

### **Preliminary Recommendation:**

The IRD-WG recommends the following requirement for a Whois service in the IDN environment:

- a) 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;
- b) WHOIS clients must be able display result of queries in both U- and A-label for the domain names; and
- c) Whois responses should include variants of an IDN label in the response as well.

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: 测试.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 above service requirement. 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).**

## Deficiencies of the WHOIS protocol:

There are increasing community concerns that the current WHOIS protocol does not meet the community's current needs. These are noted in recent reports from ICANN's Security and Stability Advisory Committee (SSAC) [SSAC 2003, 2008a, 2008b, 2009a] ], in reports of other ICANN supporting organizations and advisory committees [GNSO 2010] and by external sources [Newton 2006]. With respect to internationalization, the deficiency is described by the protocol 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."

## Query and display of Variants:

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 囯際 is also acceptable [Yao 2007].

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: 清華大學、清華大学、清華大學、清华大学、清华大學、清華大学、清華大學 [Yao, 2007].

The WG deliberated the issues of how to query and display variants extensively, and has made 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.
- WG members noted that it is outside the scope of the IRD WG to define variant or discuss how different languages handle variants. Rather, the WG will use the categories as defined (activated vs. reserved) and make recommendations.
- The WG has agreed that a Whois query of an activated variant should return the WHOIS result of the domain that it is a variant of, as well as an indication that the label queried is a variant of the original domain.
- The WG has also agreed that what a WHOIS query of a reserved variants returns is a matter of local policy. The WG has identified two options: A query of a reserved variant for XYZ domain would return a message saying that this is a reserved variant of XYZ domain or a query of a reserved variant would return the same information as the query for an activated variant. Furthermore the WG thinks it would be helpful that in the Whois result, it can provide a link to the registrar/registries' variant policy.

## What is needed from internationalized registration data to accommodate users who want to submit and have registration data displayed in “familiar” characters from local scripts?

WG members agreed that various elements of registration data (see background on Whois data) could be separately internationalized.

**Domain names (RAA 3.3.1.1):** Per requirement from the previous section, WHOIS services should return both A-label and U-label representation for the given IDN domains queried.

**Nameserver names (RAA 3.3.1.2):** Currently all nameservers are in US-ASCII. However, with internationalized domain names, it is possible that some will publish their nameservers in IDN. There are several alternatives: one way is to always display it in US-ASCII 7 using the A-label, as this information is generally only of technical interest and should be displayed in same way as it is in the DNS; another way is to have name servers displayed in both A-label and U-label to the extent such information is available. The working group felt that this field should be 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):** 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. Note that ICANN's Registrar Accreditation Application requires applicants to submit "a transliteration of the Applicant's name in Latin characters." (Need reference)

**Telephone/Fax (RAA 3.3.1.7,8):** Some IRD-WG members thought that the UPU E. 123 standard would be a good candidate 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 ongoing, some IRD-WG members thought that the email address field should be display according to RFC 5335 now and later according to the new standards when those are complete.

**Dates (RAA 3.3.1.4,5)** include 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 registration. For example client-hold, delete prohibited, update prohibited, etc. There options: 1) leave it in ASCII7; 2) always 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) could be any combination of these approaches. The WG deliberated 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 applicant guide book (WHOIS specification).

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

**Preliminary Recommendation:**

- The WHOIS data fields should be separately internationalized. Specifically, WHOIS output for domain names, sponsoring registrar, telephone/fax, and email address should be internationalized as discussed above.

**Harmonizing the labels:**

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

Regarding this point, some Working Group members suggested:

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

**Should WHOIS support multiple representations of the same registration data in different languages or scripts? Is it desirable to adopt a “must be present” representation of data, in conjunction with local character set 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.

## Description of Model 1:

Model 1 requires registrants to provide their Whois data in a “must be present” script, currently in US-ASCII. Optionally, the registrars may also ask registrants to provide their contact information in a local language as well. If registrants also provide information in their local language, then this information should be displayed. The following figure 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 (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 characters. The registrars display it in US-ASCII7.**

## Description of Model 2:

In this model, 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.

```
$ whois -h idnwhois.registrarX.ru      жук.рф
$ whois -h idnwhois.registrarX.ru  XN--F1A10A.XN--P1AI

% Registrar X WHOIS server
% This query returned 1 object

domain:      жук.рф
domain-ace:   XN--F1A10A.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 Иванов
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 (POC) to respond to translation issues.

### Description of Model 3:

In model 3, registrants provide their registration data in [any script](#) accepted by the registrar, and registrar provides tools to assist the registrant so it can be published it in a “must be present” script.

```
$ whois -h idnwhois.registrarX.ru      жук.рф
$ whois -h idnwhois.registrarX.ru  XN--F1A10A.XN--P1AI

% Registrar X WHOIS server
% This query returned 1 object

domain:      жук.рф
domain-ace:   XN--F1A10A.XN--P1AI
contact:      Petr Ivanov
organisation: OAO «Tsirkul»
address:      Office 1, Ulitsa Lenina, Kovrov
address:      Vladimirskaya oblast, 601900
address:      Rossiya
phone:        +7 49232 48720
fax-no:       +7 49232 48722
e-mail:       cicle@cicle.ru
```

**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.**

#### Description of Model 4:

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.

The WG then discussed the impact of each of these models to registrars, registries, registrants, users of Whois, the discussion is summarized in the table below.

Model 1:US-ASCII mandatory, local  
language optional

Mode 2: Any language  
accepted and displayed in  
Whois, registrar (registry)  
provides point of contact.

Model 3, 4: Any language  
accepted,  
translation/transliteration is done.

## Impact to registrars

to existing registrars	Currently ICANN accredited registrars use US ASCII-7 as the default script for Whois. 1) If they accept additional language inputs, this is additional cost to store the data, update the registration process. 2) We may want to specify in the cases of foreign registration, the contact information must be in English?[Question: Did any in our group suggest that registration not be accepted in non-English languages?] Barriers of entry as they need to support the additional language.	Under this model, registrars will provide point of contact to deal with transliteration issues for each language they accept registration in. 1) This increases the registrars cost, 2) It is also important to set some service guarantees for this point of contact, otherwise a transliteration request could be delayed.	1) Increase in the cost for registrars by providing the translation/transliteration service. 2) Uncertainty: When there is a mistake in translation/transliteration that results in inaccuracies, who is responsible? 3) Who is responsible to maintain the accuracy of the transliteration / translation?
to new IDN based registrars	Registrars need put multi languages pages on the front end. 2) If they are to require local language in addition to English, they are likely asking registrants to provide those information.	same as above	Same as above

## Impact to registries

to Thin registries	No impact, as only US-ASCII7 is submitted.	No impact	No impact
to Thick registries	No impact	The registry will not be able to interpret the registrant information unless that have a service (internal or otherwise) that can translate the script/language used. This will prevent them from engaging in 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	No impact



ccTLDs there is no official role for the registry to do any of these but if such a role were to develop then this would prevent it.

<b>Impact to registrant</b>	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 of Whois.	Least barrier of entry to registrants.	Least barrier of entry to registrants.
<b>Impact to users of Whois</b>	Model 1	Model 2	Model 3
IRD-only user	May be enhanced usability of Whois as it is in his or her own script	may be enhanced usability of Whois as it is in his or her own script	enhanced usability of Whois as it is in his or her own script
ASCII-only capable user	Unchanged, because there will be an "must be present" script	Pose significant challenges as Whois now in many scripts that the local user would not understand	Unchanged, because there will be an "must be present" scripts
to legitimate automation user	little impact	Little impact as the encoding would likely be UTF-8	little impact
<b>Technical Impact</b>			
<b>Impact to applications</b>			

Port 43 Whois clients	The WHOIS protocol has no mechanism for indicating the character set in use. Absence of protocol solution, some Whois servers that support IRD require flags to specify output encodings (e.g. .DE, .JP WHOIS servers). Maybe it is possible to specify input encodings (not sure). 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).	Not sure	Not sure
Web Whois clients	This depends on both the operating system and the browser. For characters to be displayed properly, the browser such as Mozilla Firefox or IE 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, such as Windows 95, provide partial support for Unicode.	Not sure	Not sure
Other whois applications	Not sure	Not sure	Not sure
<b>Impact to Whois system itself</b>	Not sure	Not sure	No impact

WG's made the following observations:

**On Model 1:**

- Many WG members felt both model 1 and model 2 are feasible.
- Some WG members felt that option 1 is the one that seems to have the least amount of impact potentially to registrars and registries. But in so doing it may also have the least amount of use because local language display is purely optional. So Model 1 is the least common denominator option.

**On Model 2:**

- Many WG members felt model 2 are feasible as well.
- However, some WG members raised the question that whether there will be increased *inaccuracy* due to this model. As what will probably happen is that registries will not perform any validation or verification of the validity of even the scripts of the languages coming in, and not be responsible for the accuracy of the information. They will accept whatever information is given by the registrar and enter that information in the Whois record. No verification of accuracy or a single script will be performed, so it is possible that an entry that combines Cyrillic, simplified Chinese, Hindi could be created as valid Whois entry.
- Some WG members were wary of any solution that relies upon registrar provision of a PoC, whether to the public or to registrants. ICANN has shown limited institutional capability to enforce basic compliance, let alone qualify control in similar situations.

**On model 3 and 4:**

- Many WG members raised concern to model 3 and 4 because it is costly to registrars, that the service fee for translation in many countries is very expensive, sometimes higher than the price of the domain name.
- Some felt that model 3 is not very helpful as transliteration is not precise enough to benefit law enforcement or intellectual property enforcements.
- Also some WG members felt that models 3 and 4 effectively describes 'added value'. Our focus (as with all policy groups) should be on baseline behaviour as there is an infinite spectrum of added value that we cannot predict.

- Some WG members were wary of any solution that relies upon registrar provision of a POC, whether to the public or to registrants. ICANN has shown limited institutional capability to enforce basic compliance, let alone quality control, in similar situations.

## **5** Next Steps

The Working Group welcomes the community's input on this interim report, and will use the input as well as the deliberations in the Working Group to reach a set of recommendations for internationalising Whois.

## References

1. Daigle, L. (2004) WHOIS Protocol Specification, RFC 3912.
2. Faltstrom, P., Hoffman, P., and Costello, A. (2003) Internationalizing Domain Names in Applications, RFC 3490
3. Hoffman, P. and Blanchet, M. (2002) "Preparation of Internationalized Strings", RFC 3454.
4. Hoffman, P. and Blanchet, M. (2003) "Nameprep: A Stringprep Profile for Internationalized Domain Names", RFC 3491.
5. Costello, A. (2003) "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications", RFC 3492
6. Internet Corporation for Assigned Names and Numbers (ICANN). (2007) IDNs Glossary. Marina Del Rey, CA: ICANN. Retrieved August 10, 2010, from <http://www.icann.org/en/topics/idn/idn-glossary.htm>
7. ICANN Generic Names Supporting Organization (GNSO). (2009d) *Internationalized Registration Data Working Group Draft Charter*. Marina Del Rey, CA: ICANN. Retrieved February 10, 2010, from <http://gnso.icann.org/issues/ird/ird-wg-charter-24sep09.htm>
8. ICANN Generic Names Supporting Organization (GNSO). (2010) *Inventory of Whois Service Requirement Final Report*. Marina Del Rey, CA: ICANN. Retrieved October 21, 2010, from <http://gnso.icann.org/issues/whois/whois-service-requirements-draft-final-report-31may10-en.pdf>
9. ICANN Security and Stability Advisory Committee (SSAC). (2003) *WHOIS Recommendation of the Security and Stability Advisory Committee* (SSAC publication No. 003). Retrieved from <http://www.icann.org/en/committees/security/sac003.pdf>
10. ICANN Security and Stability Advisory Committee (SSAC). (2008a) *SSAC Comment to GNSO regarding WHOIS studies* (SSAC publication No. 027). Retrieved from <http://www.icann.org/en/committees/security/sac027.pdf>
11. ICANN Security and Stability Advisory Committee (SSAC). (2008b) *Domain Name Registration Information and Directory Services* (SSAC publication No. 033). Retrieved from <http://www.icann.org/en/committees/security/sac033.pdf>
12. ICANN Security and Stability Advisory Committee (SSAC). (2009a) *Display and usage of Internationalized Registration Data: Support for characters from local languages or scripts* (SSAC publication No. 037). Retrieved from <http://www.icann.org/en/committees/security/sac037.pdf>
13. Newton, A. and M. Sanz (2005a). "IRIS: The Internet Registry Information Service (IRIS) Core Protocol", RFC 3981, January 2005.
14. Newton, A. and M. Sanz (2005b) "IRIS: A Domain Registry (dreg) Type for the Internet Registry Information Service (IRIS)", RFC 3982, January 2005.
15. Newton, A. (2006) *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
16. Yao, Jiankang (2007). RFC 3743 and IDN TLD tests. CNNIC. Retrieved August 30, 2010 from <http://losangeles2007.icann.org/files/losangeles/ChinaonIDNs.pdf>

## Annex 1 – Board resolution to create IRD Working group

*Excerpt from Board Resolutions 26 June 2009.* Marina Del Rey, CA: ICANN. Retrieved August 25, 2010, from <http://www.icann.org/en/minutes/resolutions-26jun09.htm>.

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

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.\_\_), 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.