Display and usage of Internationalized Registration Data

Dave Piscitello

SAC 037: Display and usage of Internationalized Registration Data

Reference document can be found at

http://www.icann.org/committees/security/sac037.pdf

The Problem Space

- Internet applications must accommodate a global audience using characters in local languages and scripts
- WHOIS is an Internet application, therefore
- WHOIS must accommodate a global audience

More than IDN

- IDN standards and guidelines define how Domain Names will be composed and displayed
  - WHOIS application and web developers can apply these for submission and display of domain names
- No standards or guidelines define how Domain Registration Data are composed and displayed
  - Sponsoring registrar
  - Contact information,
  - DNS configuration data…

Is this a problem or an evolution?

- The "Internationalized registration data" problem is not new…
  and not due to IDN TLDs
- IDN labels can be composed and registered at the second and lower level in many TLDs
- The user experience with WHOIS today varies

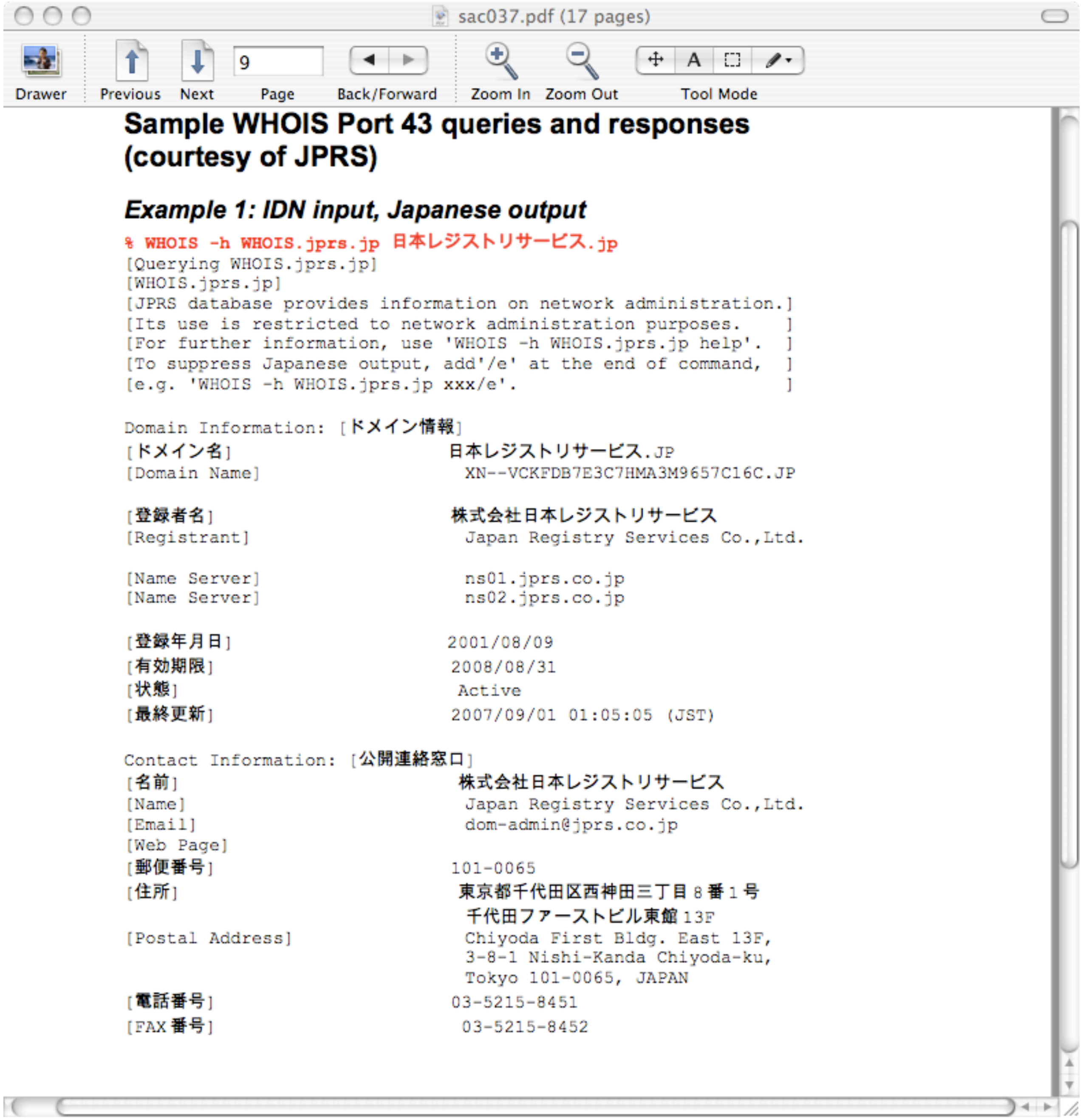
WHOIS treatment of IDN labels

- User experience in the field varies
- Internet users might attempt to submit A-labels or U-labels to query WHOIS services
  - over Port 43 using OS command line programs and WHOIS client applications
  - as input to web-based WHOIS query and domain name registration submission forms
  - WHOIS operators might display domain labels in A-label encoding or in characters from the local script used by the registrant

WHOIS treatment of Internationalized registration data

- Recognizable display of Registration Data is application dependent
- Display may be affected by character sets available to WHOIS clients and browsers
- Registrars and 3rd party WHOIS service providers may choose encodings that best represent the languages and scripts of the domain registration data

SAMPLE PORT 43 WEB OUTPUT

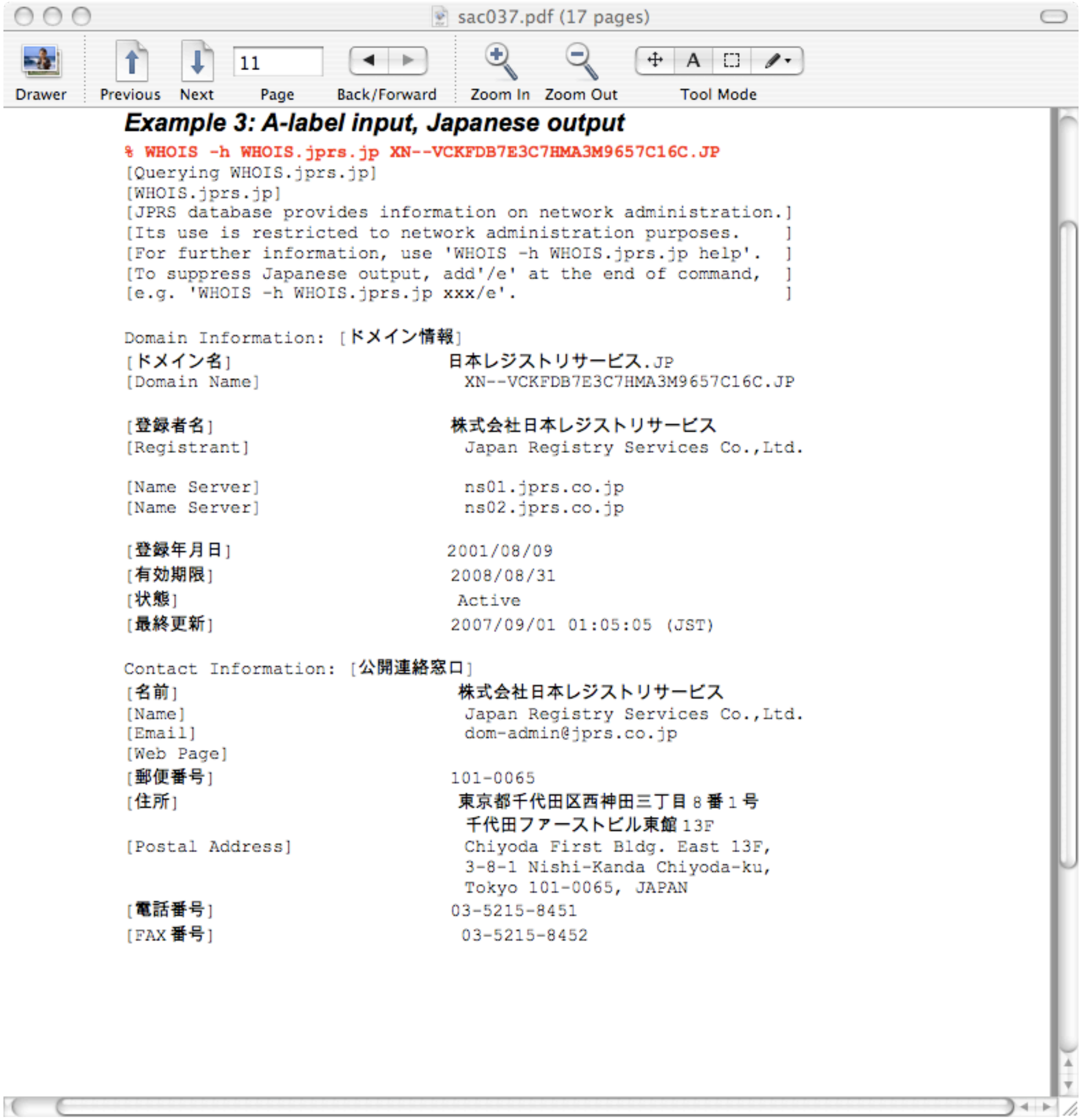


Sample WHOIS Port 43 queries and responses (courtesy of JPRS)

Example 1: IDN input, Japanese output

```
% WHOIS -h WHOIS.jprs.jp 日本レジストリサービス.jp
[Querying WHOIS.jprs.jp]
[WHOIS.jprs.jp]
[JPRS database provides information on network administration.]
[Its use is restricted to network administration purposes.    ]
[For further information, use 'WHOIS -h WHOIS.jprs.jp help'.  ]
[To suppress Japanese output, add'/e' at the end of command,  ]
[e.g. 'WHOIS -h WHOIS.jprs.jp xxx/e'.                         ]

Domain Information: [ドメイン情報]
[ドメイン名]                     日本レジストリサービス.JP
[Domain Name]                    XN--VCKFDB7E3C7HMA3M9657C16C.JP

[登録者名]                       株式会社日本レジストリサービス
[Registrant]                     Japan Registry Services Co.,Ltd.

[Name Server]                    ns01.jprs.co.jp
[Name Server]                    ns02.jprs.co.jp

[登録年月日]                     2001/08/09
[有効期限]                       2008/08/31
[状態]                           Active
[最終更新]                       2007/09/01 01:05:05 (JST)

Contact Information: [公開連絡窓口]
[名前]                           株式会社日本レジストリサービス
[Name]                           Japan Registry Services Co.,Ltd.
[Email]                          dom-admin@jprs.co.jp
[Web Page]
[郵便番号]                       101-0065
[住所]                           東京都千代田区西神田三丁目8番1号
                                 千代田ファーストビル東館13F
[Postal Address]                 Chiyoda First Bldg. East 13F,
                                 3-8-1 Nishi-Kanda Chiyoda-ku,
                                 Tokyo 101-0065, JAPAN
[電話番号]                       03-5215-8451
[FAX番号]                        03-5215-8452
```

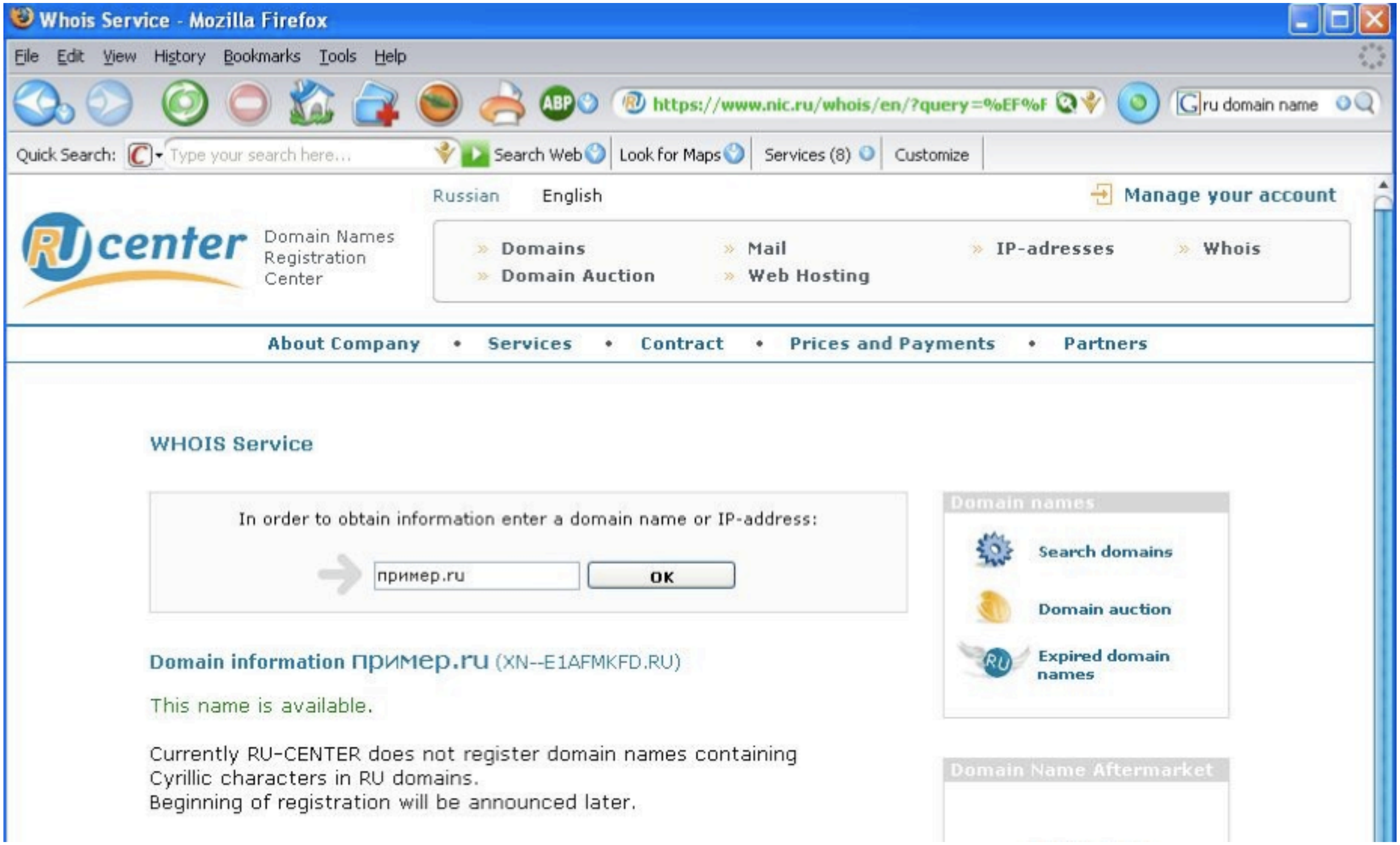
SAMPLE PORT 43 WEB OUTPUT



Example 3: A-label input, Japanese output

```
% WHOIS -h WHOIS.jprs.jp XN--VCKFDB7E3C7HMA3M9657C16C.JP
[Querying WHOIS.jprs.jp]
[WHOIS.jprs.jp]
[JPRS database provides information on network administration.]
[Its use is restricted to network administration purposes.     ]
[For further information, use 'WHOIS -h WHOIS.jprs.jp help'.   ]
[To suppress Japanese output, add'/e' at the end of command,   ]
[e.g. 'WHOIS -h WHOIS.jprs.jp xxx/e'.                          ]

Domain Information: [ドメイン情報]
[ドメイン名]                    日本レジストリサービス.JP
[Domain Name]                    XN--VCKFDB7E3C7HMA3M9657C16C.JP

[登録者名]                      株式会社日本レジストリサービス
[Registrant]                     Japan Registry Services Co.,Ltd.

[Name Server]                    ns01.jprs.co.jp
[Name Server]                    ns02.jprs.co.jp

[登録年月日]                    2001/08/09
[有効期限]                      2008/08/31
[状態]                          Active
[最終更新]                      2007/09/01 01:05:05 (JST)

Contact Information: [公開連絡窓口]
[名前]                          株式会社日本レジストリサービス
[Name]                           Japan Registry Services Co.,Ltd.
[Email]                          dom-admin@jprs.co.jp
[Web Page]
[郵便番号]                      101-0065
[住所]                          東京都千代田区西神田三丁目8番1号
                                 千代田ファーストビル東館13F
[Postal Address]                 Chiyoda First Bldg. East 13F,
                                 3-8-1 Nishi-Kanda Chiyoda-ku,
                                 Tokyo 101-0065, JAPAN
[電話番号]                      03-5215-8451
[FAX番号]                       03-5215-8452
```

SAMPLE WHOIS WEB OUTPUT


Whois Service - Mozilla Firefox
File
Edit
View
History
Bookmarks
Tools
Help
https://www.nic.ru/whois/en/?query=%EF%F
ru domain name
Quick Search:
Type your search here...
Search Web
Look for Maps
Services (8)
Customize
Russian
English
Manage your account
RU center
Domain Names Registration Center
Domains
Domain Auction
Mail
Web Hosting
IP-adresses
Whois
About Company
Services
Contract
Prices and Payments
Partners
OK
Domain names
Search domains
Domain auction
Expired domain names
Domain Name Aftermarket

WHOIS Service

In order to obtain information enter a domain name or IP-address:

пример.ru

Domain information пример.ru (XN--E1AFMKFD.RU)

This name is available.

Currently RU-CENTER does not register domain names containing Cyrillic characters in RU domains.
Beginning of registration will be announced later.

Questions to Consider: User Experience Issues

- What features will Internet users find most beneficial in applications that bind registered domain names to Registration Data when the Internet user experience becomes more diverse with respect to languages and scripts?
- Are there any general principles that registry operators and registrars could adopt to minimize the "Babel effect" on Registration Data query services and to ensure some uniformity of information display?

Questions to Consider:
Data Reliability, Accuracy, Operational Issues

- Are there operational issues related to the submission and display of Registration Data in Unicode in non-Web mediums?
  - Are applications that automate DNS and Whois queries affected?
- How are parties affected when a Registration Data query returns some or all non US-ASCII characters in its output?
  - Impact on law enforcement, brand protection, analysis of malicious conduct
- What effect has the use of non US-ASCII characters in Registration Data already had on domain registration data accuracy programs?

Questions to Consider: Security & Standardization Issues

- What information, and in what languages and scripts, should be permitted when collecting and displaying registration data for a (set of) domain name(s)?
- Do sufficient submission and display practices exist for applications that bind registered domain names to information about Registration Data?
- Should the maintenance and display certain registration data be required in US- ASCII to ensure a common denominator for core information display

SSAC Recommendations

- ICANN's Board of Directors should task the GNSO, ALAC, CCNSO, GAC and SSAC to form a WG to study the feasibility and suitability of introducing display specifications to deal with the internationalization of Registration Data.
  - Representation from ccTLD operators and consultation with the CCNSO should be a part of the Working Group's charter.

Feasibility Study

- Consider feasibility of having applications that query Registration data incorporate "standard" functionality
- Example: require application to bind registered domain names to registration data as follows:
  - ✧ accept A-label and U-label domain names as input
  - ✧ return A-label and U-label domain names as output
  - ✧ store contact information in XML
  - ✧ retain Unicode encoding of local characters in stored contact information
  - ✧ accept Record and Key Type data queries in Unicode
  - ✧ return responses to record and key type data queries in Unicode
  - ✧ return responses including character or word variants bundled with each Registration Data query

Preparing for ICANN Seoul Meeting

- SSAC members and ICANN staff looking at
  - Current CCTLD WHOIS submission and display practices
  - Other conventions and standards for internationalizing contact information
  - Richer data schema for domain registration records
- ICANN Staff, SSAC, GNSO preparing draft
  - Working Group charter and implementation plan

Should the maintenance and display of certain registration data be required in US- ASCII to ensure a common denominator for core information display?

Staff & SSAC member study/exercise

What do current WHOIS practices tell us?

- Solicited current practices from 16 CCTLDs by asking:
  - ✧Does your registry allow users to register domain names using characters from local scripts?
  - ✧Does your registry collect and store registration data in US-ASCII7 in addition to characters from local scripts?
  - ✧Can users of the web interface choose the display language? What languages does your web interface support?
  - ✧Does your registry provide access to registration information via WHOIS/port 43? Can users choose the display language?

Current WHOIS practices (responses)

- 10 0f 16 CCTLDs registry allow users to register domain names using characters from local scripts
- 10 of 16 CCTLDs support “English” (US-ASCII7) and a local language/script, including
  - Arabic, Chinese, German, Japanese, Lithuanian, Portuguese, Spanish, Swedish
- All CCTLDs support WHOIS/Port 43
- Character set dependencies affect WHOIS client dependencies submission and display
  - Must support UTF-8, UTF-16, or ISO-8859

Are there any general principles that registry operators and registrars could adopt to minimize the "Babel effect" on Registration Data query services and to ensure some uniformity of information display?

Staff and SSAC member study/exercise

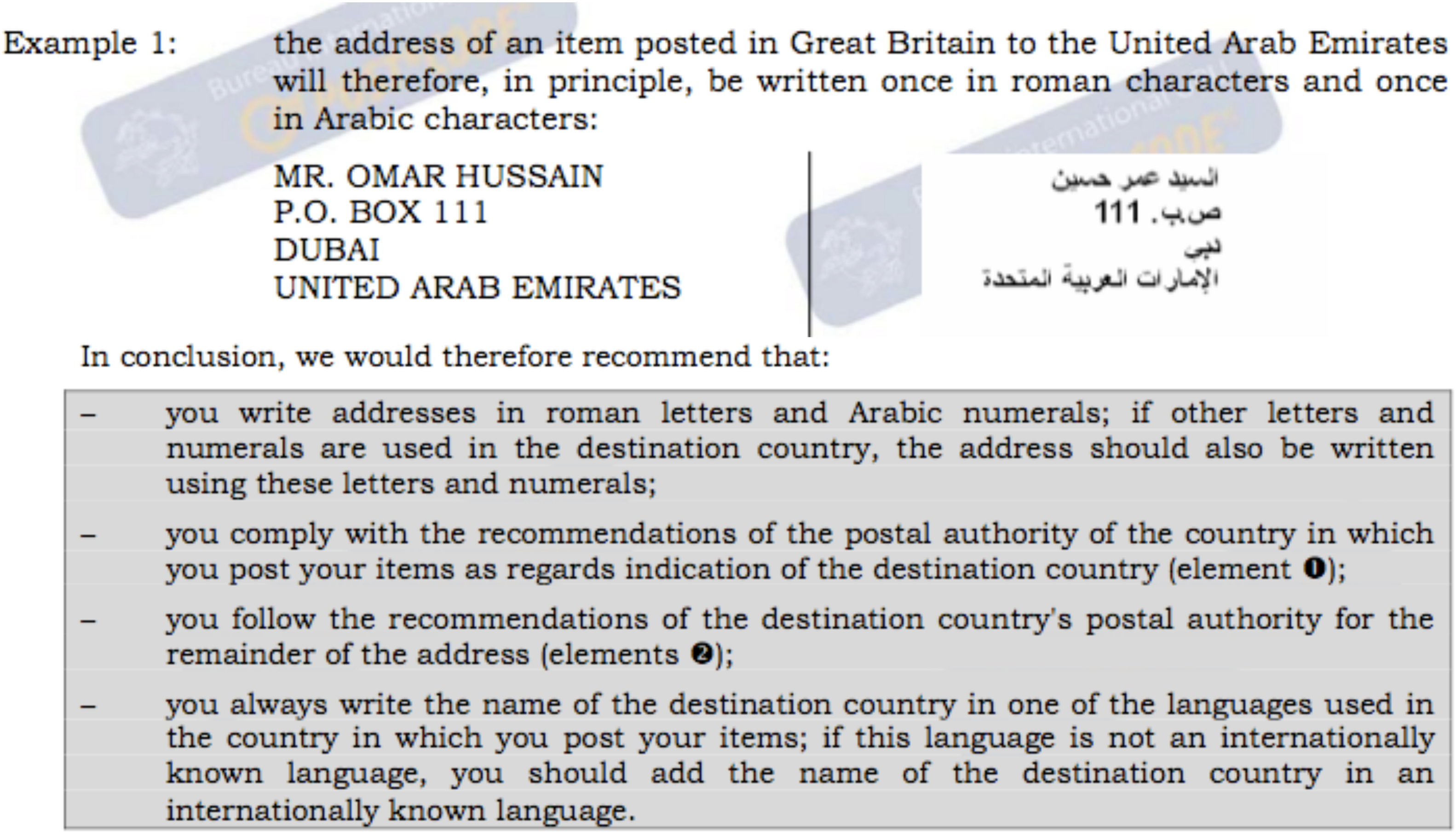
Other conventions and standards

- Universal Postal Union, International Address Formatting

  "The addressee's address shall written legibly in roman letters and Arabic numerals.

  “If other letters and figures are used in the country of destination, it shall be recommended that the address be given also in these letters and numerals. “
- Question: Is “must US-ASCII7, also <local script>” appropriate for registration data?

Other conventions and standards



Example 1: the address of an item posted in Great Britain to the United Arab Emirates will therefore, in principle, be written once in roman characters and once in Arabic characters:

MR. OMAR HUSSAIN
P.O. BOX 111
DUBAI
UNITED ARAB EMIRATES

السيد عمر حسين
ص.ب. 111
دبي
الإمارات العربية المتحدة

In conclusion, we would therefore recommend that:

- you write addresses in roman letters and Arabic numerals; if other letters and numerals are used in the destination country, the address should also be written using these letters and numerals;
- you comply with the recommendations of the postal authority of the country in which you post your items as regards indication of the destination country (element ❶);
- you follow the recommendations of the destination country's postal authority for the remainder of the address (elements ❷);
- you always write the name of the destination country in one of the languages used in the country in which you post your items; if this language is not an internationally known language, you should add the name of the destination country in an internationally known language.

http://www.upu.int/post_code/en/formatting_an_international_address_en.pdf

What information, and in what languages and scripts, should be permitted when collecting and displaying registration data for a (set of) domain name(s)?

Data schema for successor to WHOIS protocol

- SAC027, SAC033, and SAC037 all encourage ICANN community to consider a successor to the WHOIS protocol that can improve accuracy and availability of domain registration data
- Staff and SSAC studying IRIS/CRISP/DREG standards and prior WHOIS studies to inventory service requirements
- Before enumerating requirements, taking stock of the data we collect to understand what data need to be added to support additional/future services

Questions?