

Generic IDNs and the New gTLD Program

For those TLDs which do not represent a country name, non-ASCII scripts will be introduced at the same time as the ASCII new generic domains (e.g. the equivalents of .com in simplified Chinese or .car in Kanji). This introduction will be subject to the New gTLD Program. After approval from the ICANN board this program was launched in January 2012.¹ The New gTLD Program focuses on offering a wider span of choice for generic top-level domains, or gTLDs. This expansion includes IDNs at the top level and is required to meet growing diversity and encourage competition for more innovation, choice and change to the Internet's addressing system. ICANN has developed both a process for registries to apply for new gTLDs and an Applicant Guidebook that takes applicants through the process and explains the implications of the many complicated issues surrounding these new top-level domains.

The Future

While the exact date for the introduction of new gTLDs in the root is unclear, it is expected that by autumn 2013 the number of TLDs (currently standing at just over 300) will have increased significantly. An important part of that increase can be attributed to the introduction of IDN TLDs. At the time of writing 38 IDN ccTLDs (in addition to the 6 test TLDs) have been added to the root zone, and more will be added soon. As the introduction of the Cyrillic .рф has demonstrated, one can expect that a well managed IDN TLD will succeed in unleashing the enormous potential of those user communities that have so far been deprived of convenient access to the World Wide Web.



SUMMARY

This paper aims at introducing the reader to Internationalized Domain Names (IDNs) and providing an overview of the reasons why the implementation of IDNs was necessary. It also explains the different processes that lead to their introduction in the Domain Name System (DNS) root zone and provides an overview of the current situation.

INTERNATIONALIZED-DOMAIN-NAMES:WHAT-&-WHY?.01

At the time of the introduction of the Domain Name System, designers of the DNS (and its predecessor, the "host.txt" file) wanted to allow for non-ASCII¹ characters to be used within the system, but the technology in use at that time was simply not powerful enough to accommodate this². Additionally, the group of users was well defined and restricted: users of ARPANET and its successor were, well into the eighties, mainly US academics or research institutions. Even with the increasing internationalization of the web, those users had one thing in common: they used English to communicate and therefore had no other needs than the ASCII characters (Basic Latin script, Arabic numbers 0-9 and the hyphen) to create and use these humanly meaningful addresses (or domain names).

With the global deployment of the Internet and the exponentially increasing user base, English was still used as the Lingua Franca, but it became clear that the technical restriction which limited the characters that can be used in one script became a significant obstacle for large communities of users in e.g. the Arabic region, China, Indonesia or India. This limitation made it very difficult or even impossible for those users to connect with and interact over the web.

In response to such technical restrictions, these communities developed mechanisms that could partially overcome this hurdle by introducing hybrid domain names. While the root zone still only held ASCII-based top level domains, some of the registries operating those top level domains introduced the possibility of making use of different scripts in the second and third levels. Internationalized Domain Names are domain names that include or consist of different scripts such as Cyrillic, Hangul or Arabic.

This solution was however not regarded as satisfactory as it indeed led to a number of problems³.

- Complexity of typing: Hybrid names still required the user to switch keyboards when typing a domain name
- Confusion over label order: As some scripts are right-to-left, the direction could switch in a domain name.
- Ambiguity of visual appearance of different domain names: Different domain names could look identical depending on the input mechanisms of the application (e.g. browser).

The technical community – united within the framework of the Internet Engineering Task Force – initiated work on standards for Internationalized Domain Names in Applications (IDNA) in 2003. These standards⁴ provide technical guidelines for the

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

2 Internationalization of Domain Names: a history of technology development.; Klensin, J. and Fälström, P.; <http://www.isoc.org/internet/issues/docs/i18n-dns-chronology.pdf>

3 A detailed overview of these difficulties can be found in "Internationalized Domain Names: The Long and Winding Road"; Benny Lipsicas, Doron Shikmoni; Domain Wire 2007; p7-11 https://www.centr.org/system/files/agenda/attachment/domainwire_2007.pdf

4 These are RFCs 3454, 3490, 3491, and 3492 – now obsolete by the introduction of RFCs 5890 and 5891

1 <http://singapore41.icann.org/meetings/singapore2011/transcript-board-new-gTlds-20jun11-en.txt>

deployment of IDNs and describe a translation mechanism able to translate any standard script (or more precisely its Unicode equivalent) into a valid DNS character set. It should be noted that these technical guidelines are currently only applicable to the Domain Name System and do not provide a solution for other protocols¹. With the additional introduction of ICANN's guidelines for IDN implementation² on the second level, the technical requirements were in place for the full introduction of non-ASCII characters in the DNS. In April 2007, the ICANN Board endorsed version 2.2 of the implementation guidelines which made these guidelines also applicable to the top level³

THE-DIFFICULT-ROAD-TO-IDNs-IN-THE-ROOT.02

During initial discussions within the ICANN community, it became evident that full deployment of IDNs in the root zone would take much too long to accommodate the immediate needs of the communities requiring these. While the technical standards were in place, the policy and political aspects of the introduction of IDNs were still quite complex and subject to long debates within the ICANN community. To name just a few:

- Which registries would be allowed to manage the new top level domains?
- Would the name of country X in a script that is not used in country X be assigned to country Y that is using that script? (E.g. would "Sweden" in Chinese be managed by a Swedish registry or by a Chinese registry?)
- How could political struggles between countries or territories be avoided for identical names?
- How could confusion be avoided between names that looked similar to the end-user even though they were written in different scripts⁴?
- How could it be possible to avoid undermining accessibility of the Whois tool through the use of dozens of different scripts?

The final push for the introduction of IDNs came from the imminent threat that the Internet could break up in different zones, handling different scripts. It was expected that in particular the Russian Federation, the Arabic region and China would not continue to endlessly wait until their scripts were allowed in the DNS. This would have led to a scattered DNS and could have signaled the end of the Internet as we know it.

Answering calls from the Government Advisory Committee (GAC) and the country code Name Supporting Organisation (ccNSO), the ICANN Board agreed to split the introduction into two phases: a fast-track process with limited scope to fulfill the near-term demand and a regular Policy Development Process (PDP) that meets the long-term demand of those that did not qualify for the fast-track process⁵. The introduction of generic IDN names (such as .car or .com in non-latin scripts) will take place under the New gTLD Program.

IDN ccTLD Fast Track process

The ICANN Board approved the implementation plan on October 30th 2009. The process was launched on November 16th 2009. The application was conditional upon fulfilling the following criteria:

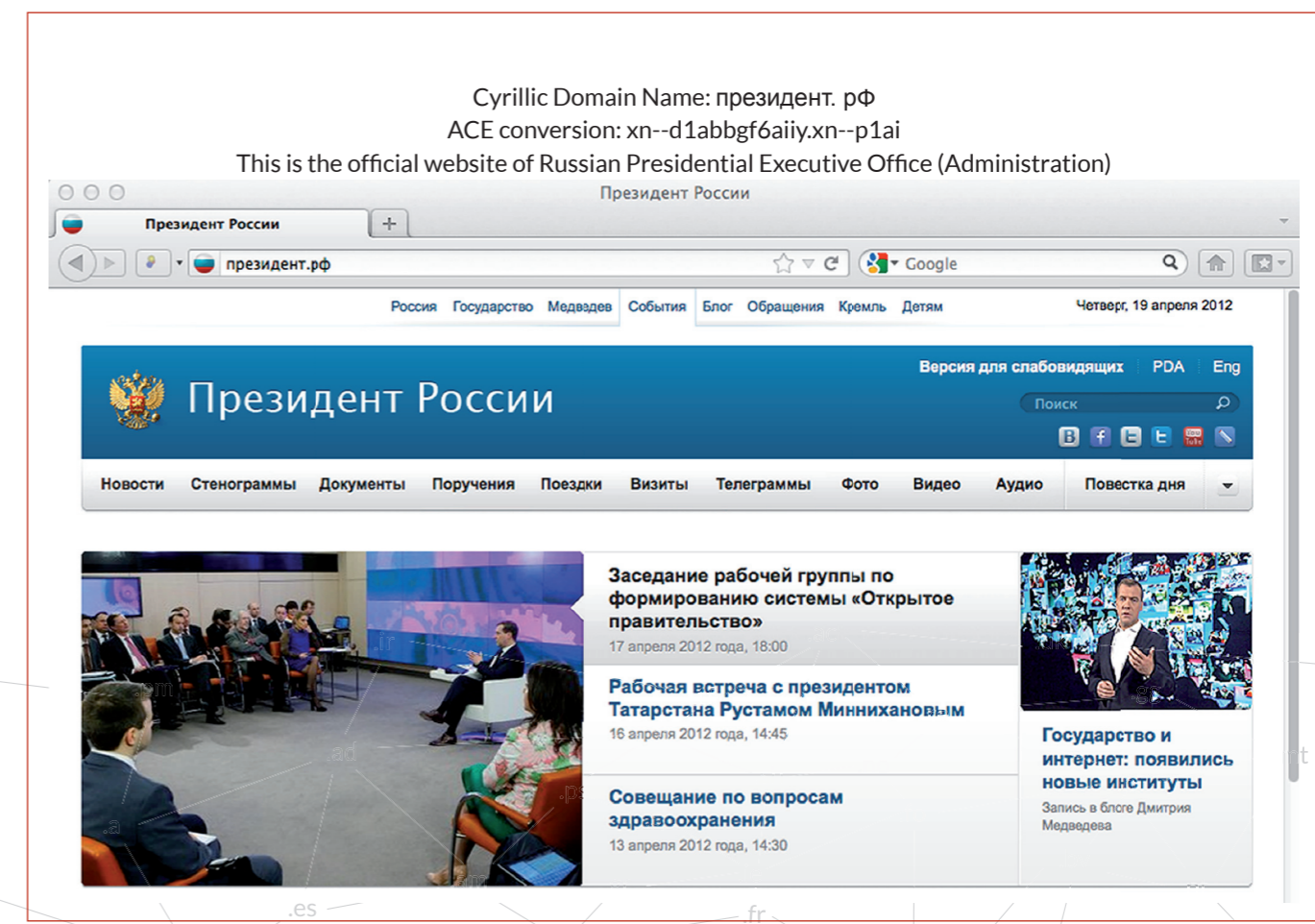
1. Access to the Fast Track process is restricted to countries or territories that appear in the ISO 3166-1 list⁶.
2. Every application needs to be accompanied by demonstrated community support
3. Applied for strings need to be meaningful representations of the corresponding country or territory name
4. Applied for strings should not be confusingly similar to existing strings⁷

A striking example of the need for IDN ccTLD names is the very successful introduction of the Cyrillic equivalent to .RU - .рф. Since its launch on 11 November 2010, more than 790.000 domain names have been registered. Compared to the 3 million domains under .RU (operational since 1994) this demonstrates the significant potential to include users that have previously been left out.

1 IDN usage in emails—or, more specifically, in the domain name part of the email address—is not yet fully available. The technical standard that will make this possible is being developed by the Internet Engineering Task Force.
 2 These guidelines are applicable to second level domains. <http://icann.org/topics/idn/implementation-guidelines.htm>
 3 Version 2.2 of the ICANN IDN implementation guidelines: <http://www.icann.org/topics/idn/idn-guidelines-26apr07.pdf>
 4 The standard example is "www.paypal.рф" which is a full Cyrillic domain name but looks identical to the hybrid "www.paypal.рф". This would make the system vulnerable to fraud
 5 A full overview of the Fast Track process is available at: <http://www.icann.org/en/topics/idn/idn-activities-seoul-28oct09-en.pdf>
 6 This list is maintained by the International Organization for Standardization, it can be consulted at: http://www.iso.org/iso/country_codes.htm
 7 The notorious case (which was refused) is the application by Bulgaria for the Cyrillic version of .BG (.БГ) which was deemed to be confusingly similar to the Brazilian extension .BR (equivalent to .br or .Br)

By the end of January 2011 16 IDN ccTLDs had been inserted in the DNS root zone following a successful application under the Fast Track process:

- China (Simplified Chinese): .中国
- China (Traditional Chinese): .中國
- Egypt: .مصر
- Hong Kong: .香港
- Jordan: .الأردن
- Occupied Palestinian Territory: .فلسطين
- Russia: .рф
- Qatar: .قطر
- Saudi Arabia: .السعودية
- Sri Lanka (Sinhalese): .ලංකා
- Sri Lanka (Tamil): .இலங்கை
- Taiwan, province of China (Simplified Chinese): .台湾
- Taiwan, province of China (Traditional Chinese): .台灣
- Thailand: .ไทย
- Tunisia: .تونس
- United Arab Emirates: .امارات



IDN ccTLD Policy Development Process

For those country code top level domains that do not meet those criteria, the country code name supporting organization (ccNSO) is developing a policy that will allow the introduction of all country names in all scripts. This program will take due account of the lessons learned from the Fast Track process. The new policy will also address some of the organizational questions that are raised with the introduction of IDN ccTLDs (e.g. what is their voting status in the ICANN multi-stakeholder model?).