

Effect of Multilingualism on the Internet

International Issues that Affect How Governments and Economies
Address Issues Relating to a Global Infrastructure

Dr. Paul Twomey, President and Chief Executive Officer
Internet Corporation for Assigned Names and Numbers

Overview

The term “multilingualism” in the context of the Internet relates primarily to two areas: multilingual online content, and access to such content by the use of domain names that include non-ASCII characters (called internationalised domain names, or IDNs) in the Internet’s unique identifier system (in particular, the domain name system, or DNS). ICANN is responsible for the global coordination of the Internet’s system of unique identifiers. These include domain names—the names under.org and .museum and under country codes such as.uk—as well as the addresses used in a variety of Internet protocols. An example of a domain name is www.icann.org.

The IDNs are available for use at the second level of the domain name system in a multitude of non-ASCII scripts, including right-to-left scripts, Arabic being the most widespread. They also include languages based on non-alphabetic scripts, of which the largest contemporary language is Mandarin Chinese. ICANN is actively involved in the efforts to make these IDNs available at the top level—that is, so that the entire domain name is rendered in local language characters.

Currently, only about 35 percent of all Internet users are native English speakers, although English websites continue to dominate, with approximately 68 percent of all sites readable only in English. About two-thirds of English-language sites are devoted to e-commerce, and fully half of those still originate in North America. Altogether, e-commerce sites accounted for nearly US\$2.4 trillion in business worldwide in 2006. Once much higher, these numbers have gone through a natural realignment as Internet use continues to expand geographically. Thus, multilingual content is critical to the Internet’s continued evolution and use by people from all linguistic backgrounds.

Naturally, people are more comfortable reading the languages and writing the scripts they find most familiar. It follows that the promise of content on the Internet in a preferred language generates increased local interest and use. A multilingual Internet will enhance the local Internet experience in large regions of the world by enabling people to share and access information or use services offered in their own languages. A modern Internet will also be the impetus behind the growth of translation services and instant translators such as the ones provided by Google. These facilities are essential to preventing a multilingual Internet from becoming the infrastructure of a modern Tower of Babel.

Obviously, information available in different languages can enrich awareness about myriad cultures, foster appreciation of diversity, and perhaps, inspire greater respect and curiosity and decrease fear or animosity towards an unknown.

Before people around the world can enjoy this experience, however, many political, policy, cost, and technology issues remain to be resolved through collaboration among Internet stakeholders in all these realms.

Policy and Political Concerns

For multilingual content, politics and policy challenges arise in two interrelated areas. National laws and cultural norms differ on what is considered acceptable content, and this directly impacts the kinds of content that can be generated for use on the Internet. While some regions offer little governmental intervention, others have much more restrictive policies governing both content and access to such content.

In all cases, a nation may wish to raise awareness of the importance of generating information in local languages in order to encourage greater multilingual Internet content and use. For example, the government of India recently launched a countrywide campaign to encourage its more than one billion citizens to generate Internet content in many of its 22 official languages. As part of this campaign, it is providing CDs free of charge that contain instructions for generating content. UNESCO has also launched similar initiatives in many countries, as is seeing considerable success.

Both developed and developing nations are becoming increasingly reliant on the Internet as a unique and essential tool with which to communicate, to conduct business, and to store and transmit data. Prudent political and policy actions affecting this single, globally interoperable medium will continue to lead towards innovations in commerce and communication, and in our social lives.

How can multilingualism and the fact that many people worldwide do not use ASCII scripts be taken into account in the design of a future infrastructure?

Born of a 1960s-era project focused on U.S. defense, the original domain name system was based on the ASCII character sets: the 26 letters a to z, the 10 numerals 0 to 9, and the hyphen. No one could have foreseen that the domain name system would evolve into its present form or that the ASCII characters alone would be far too limited, especially in light of today's demands for global accessibility in localised scripts.

Driven by the geographic increase in the use of the Internet, the need to deploy internationalised top level names has emerged in tandem with other internationalisation efforts. At present, we are witnessing a push-pull dynamic between politics and technology in the implementation of IDNs, with policy issues adding a further layer of complications.

The governments of many countries continue to push towards the deployment of internationalised top level names to ensure continued global interoperability of the Internet's unique identifier system in certain regions. Technology developers are taking a more deliberate approach that involves rigorous testing at each level of the system to ensure operational stability and to guard against fragmentation of the Internet. The challenge facing all three of these drivers—political momentum, technological feasibility, and policy making—is to ensure that the implementation of IDNs takes place in a manner that does not jeopardise the Internet's continued stable, secure, and global interoperability. An unstable Internet serves no region's interests and could conceivably lead to fragmentation of the Internet.

The technology community has made significant gains in recent years. IDNs at the second level of a domain name have been available in test beds since 2001 and in production since 2003. Since then, ICANN has been actively involved in the efforts to move IDN deployment at the top level forward at the fastest pace possible.

In 2003, ICANN and leading IDN registries such as cn, info, jp, org, and tw announced the commencement of global deployment of IDNs in second level domains.

Early in 2004, ICANN approved a preliminary set of procedures to improve the mechanisms for transferring IDNs between registrars. ICANN also hosted the first of many IDN workshops in Kuala Lumpur and in Cape Town, South Africa.

During 2005, ICANN released a statement regarding IDN homograph concerns and published a formal call for papers concerning deployment of internationalised top level domains. That year also saw the refinement of IDN implementation guidelines to reflect the experience gained during implementation testing. The final revision focussed on eliminating spoofing between labels; that is, the use of confusingly similar looking characters.

In 2006, ICANN and its Generic Names Supporting Organisation (GNSO) worked to develop policies for the introduction of IDNs. ICANN also engaged Autonomica AB to develop, conduct, and report on the results of laboratory testing. The laboratory technical tests of the viability of internationalised top level names and the effect they may have on the DNS are being done before insertion of such labels in the root servers. The laboratory test focuses solely on a replicate that is as close as possible to the server software of the various root servers. As such it does not incorporate the end-user perspective or a live root test.

The test design was finalised in December of 2006 and the test results are imminent. Autonomica also performed a successful feasibility test that same month.

IDN technical tests are divided into two phases, of which the first—laboratory testing to demonstrate root software and resolver stability—is well under way, as mentioned earlier. The second phase will focus on replication of the laboratory test but will be performed in the live root system.

Also in 2006, the Internet Engineering Task Force (IETF) began a review of the IDN protocol first developed in 2003. This review was initiated to enable the use of the additional characters that are entered in Unicode to be available in IDNs. In addition, the protocol is being reviewed for any necessary changes based on the implementation experience over the past few years.

The outcome of all this coordinated effort will be internationalised top level domain names that finally make entire domain names available in local scripts. However, the http:// portion will remain in ASCII, as this is the protocol prefix and cannot be changed.

What is the ICANN community's position on multilingualism?

Among ICANN's core values are preserving and enhancing the operational stability, reliability, security, and global interoperability of the Internet while respecting the creativity, innovation, and flow of information made possible by the Internet. These

goals can only be met by limiting ICANN's activities to those matters within its mission requiring or significantly benefiting from global coordination.

Thus, while ICANN's narrow mandate does not cover the content space or online content, we applaud efforts like those of the government of India to add meaningful content in local languages. We also applaud projects like those undertaken by UNESCO to promote the development of content on a world scale. We also welcome regional and national efforts to expand access to that content through the implementation of IDNs as well as through improved connectivity.

On the technical side, ICANN is committed to its responsibilities in the implementation of IDNs, including working with the technology community on the implementation at both the second level and the top level.

The ICANN community is committed to implementing IDNs in a manner that does not place the global interoperability of the Internet's unique identifier system at risk. We are convinced that before IDNs can be implemented in the root, there are many issues such as stability, intellectual property, and others that must be resolved before we can take advantage of this advance in Internet accessibility.

ICANN's current testing efforts are designed to demonstrate any negative impact that the insertion of internationalised top level domain names may have on the root server system. As mentioned earlier, the initial feasibility testing of the design of the laboratory setup was conducted in December 2006, and showed positive results. The root zone tests to follow are expected to show equal success.

Alongside these technical initiatives, ICANN's supporting organisations and advisory committees, including the Generic Names Supporting Organisation (GNSO), the Country Code Names Supporting Organisation (ccNSO), and the Governmental Advisory Committee (GAC) are working together to develop policies for the introduction of IDNs. The strict policy development protocols these organisations follow are aimed at enabling ICANN to implement a process for receiving proposals from applicants wishing to introduce new top level domains, which could be internationalised top level domains.

Conclusion

There is much work to be done as the Internet evolves, at the technical level and in resolving the regulatory, cultural, national, and social implications surrounding every innovation. The Internet's many stakeholders must work together to bridge the digital divide so that the billions of potential users now hampered by technical, practical, political, or cost considerations are assured access at the highest speeds technically feasible.

ICANN considers it important that these broader issues receive the attention they deserve in forums suited to address them. ICANN will do its part in the areas of its competence, but resolving the issues of multilingualism in terms of content development and access will require the involvement of governments, academia, the business and private sectors as well as civil society.