



Inside

- Why are IDNs important for the Asia-Pacific?
- How do IDNs relate to broader multilingual computing issues?
- Limitations of the Domain Name System
- ICANN's IDNA and DNAME testbeds
- What other systems are available for multilingual domain names?
- What are the differences of opinion on IDN implementation?
- Conclusion
- Additional Reading

APDIP e-Notes present an analytical overview of specific issues related to information and communication technologies for sustainable human development in the Asia-Pacific region. APDIP e-Notes are developed by the United Nations Development Programme's Asia-Pacific Development Information Programme (UNDP-APDIP) based at the UNDP Regional Centre in Bangkok, Thailand. For more information, visit <http://www.apdip.net> or contact info@apdip.net

This document is released under a Creative Commons Attribution 2.5 License

Summary

Internationalized Domain Names (IDNs) have become a hot topic in the field of Internet governance. As the number of non-English speakers on the Internet grows exponentially, the limitations on the Domain Name System (DNS) overseen by the Internet Corporation for Assigned Names and Numbers (ICANN) have become evident to a wider range of people. ICANN has acknowledged this with the ICANN President appointing an Advisory Committee on the issue.

The history of IDNs in the Asia-Pacific, however, goes back to testbeds established by the Asia Pacific Network Group in 1998. There are also a number of IDNs already established within particular Internet Service Providers (ISPs), with organizations such as the Multilingual Internet Names Consortium (MINC) attempting to develop a coordination framework to ensure that fragmentation of the Internet does not occur through "leakage" of these IDNs into different zones.

From the perspective of the North American/European Internet governance bodies, a single system for IDNs should be established which can serve the interests of all stakeholders, and multiple systems should be avoided. This "universal" approach to IDNs raises much more complex technical, political and economic issues than developing a viable system for a particular language group. This complexity partially accounts for the slow progress on IDN development within the ICANN system.

These are the two philosophies on IDNs: those supporting universality, standardization, stability and control on the one hand; versus multiplicity, diversity, coordination and responsiveness to local language groups on the other.

The major challenge will be to create viable mechanisms for mediating between these philosophies. The goal will be to ensure that the Internet remains a single, interoperable public facility, while ensuring that the right of all people to communicate in their own language is maintained and expanded within this new medium.

Why are IDNs important for the Asia-Pacific?

Language is the basis of communication, and use of one's own language is a basic human right. UNESCO states that language is "not only a tool for communication and knowledge but also a fundamental attribute of cultural identity and empowerment, both for the individual and the group."¹

This cultural identity and empowerment comes from seeing communication and identification occurring in a way that seems natural in one's native language. While the bulk of the content on the Internet has been in English, this is increasingly changing. In China for example, over 60 million of the nation's 100 million-plus users browse the web only in Chinese. While it is true that these users are currently able to have access while using Roman script in domain names, a truly globalized Internet would enable all users to use their own language for navigational purposes. This is especially important in areas such as education, e-government and e-commerce. Success of e-government in countries not using Roman script depends on ensuring that the ordinary citizen can remember the web address of the election website, or the e-forms and e-services which the government is offering as a public service in all the languages which are used in that country.

¹ Education in a multilingual world, UNESCO Education Position Paper No. (ED-2003/WS/2) - UNESCO, 2003, p9.

There are economic as well as socio-cultural implications to the limitations on IDNs. As economies move into information-based industries, the role of language becomes more significant. Services industries are primarily founded on knowledge and data which are expressed through language. A corporation's brands, or names, symbols, and designs that identify products or services, have an increasing amount of financial value.

During the 1990s, the DNS shifted from being a technical look-up service with little commercial impact to a key part of corporate branding strategies, with some domain names being sold for millions of dollars. Whether this value would be replicated in other languages is unknown but it remains a fact that the current DNS supports an economic market for Roman-script communities to the exclusion of others.

Realizing the value of the Internet as a truly global medium for social and economic development will require integration of communication languages that are used for offline communication and commerce. The Asia-Pacific region contains the majority of the world's languages and so multilingual domain names are particularly important for this region.

The Working Group on Internet Governance Background Report clearly identifies many of the implications of the IDN issue and areas where progress is needed, noting that:

"The current market led approach to IDN only tends to maximize the number of domain names that are sold. However, there might be cases in which global public service issues should be considered – for example, whether gTLDs² should be required to support all scripts, including minority scripts that might not be commercially viable. Without these considerations, IDN might become available only for scripts used by big countries and communities, thus contributing to the loss of linguistic diversity."³

A single DNS dominates global Internet use by agreement of major ISPs, rather than by mandate from a central body or government. If effective IDNs are not implemented within the ICANN system, it is likely that alternative navigation systems for specific language groups will continue to be developed outside of US-based coordination mechanisms such as ICANN. The differing views on how this would affect the Internet are at the core of the IDN debate.

How do IDNs relate to broader multilingual computing issues?

As with many new systems, script encoding for computers was invented for a relatively narrow set of situations with little thought for what might happen if those applications expanded rapidly. In the early days of computing within the US and Europe, there was little need to consider the use of other scripts, as computers

performed highly specialized calculation tasks, rather than the very broad range of tasks they perform today. In particular, computer systems would rarely have to interact with others, and so a diversity of standards could emerge for the encoding of characters.

Initially, the range of encoding schemes was not a significant issue, as users would generally communicate within language groups before the widespread popularity of the Internet. However, as the need for interoperability between different systems increased, Unicode emerged as an architecture that could represent any script in machine-readable code. Unicode would conceivably allow any computer to represent any character, as opposed to the script developments where language communities would simply repurpose the code points provided within operating systems designed for Roman script.

Controversies during the development of Unicode⁴ highlight the issues that continue to cause problems for IDN deployment. A single script-based encoding - rather than a language-based encoding - requires entirely new political relationships between different language groups who share that same script. In many cases, the governments have previously controlled language representation in their own territories, but in unified systems like Unicode, negotiations must take place on areas where there is overlap. Similarly, for IDNs to become universally resolvable will require political choices and compromises. While on the surface it would appear that a simple solution to DNS issues would be to allow any Unicode script encoding to be used for domain names, this is more complex than it first appears.

Limitations of the Domain Name System

The DNS was developed to solve what was, relative to today, a limited problem: how to provide a naming system more flexible than an early system called "hosts.txt" which mapped names to numerical Internet Protocol (IP) addresses. In 1984, RFC 920 established a set of top level domains (TLDs) including .com, .edu, .org, .mil and .gov to provide domain space for corporations, non-profits, schools, networks, US government offices and the US military. In these developments, assumptions were made about "the user" which would have unforeseen consequences as the Internet's reach expanded.

In some cases the constraints of the DNS have been "hacked" to permit uses broader than originally intended - for example, the country code TLD for Niue (.nu) is often put to use in Scandinavian countries where it functions as an alternate TLD. Likewise, Tuvalu's country code .tv is predominantly used for domain names associated with television. These unforeseen uses highlight the level of demand for domain names with particular mnemonic values outside the official English-derived TLDs.

² gTLDs stands for generic top-level domains, such as .com, .edu, .gov, .int, .net, .org

³ Background Report – The Working Group on Internet Governance, June 2005, p.23. <http://www.wgig.org>

⁴ An overview of debates relevant to Japanese scripts is at <http://www.jbrowse.com/text/unij.html>

The DNS was originally designed for a Roman script, and by agreement a subset of ASCII is used, referred to as "LDH": a combination of the letters a-z, digits 0-9, and hyphen. A number of issues make it difficult to upgrade the DNS to accept other scripts and languages.

To take an example, the DNS automatically maps lower to upper case - APDIP.NET is effectively the same as apdip.net. But in Canada and France there are different rules about how accents are handled when being converted between cases - which rules should the DNS use? Another example is the "a" with a dieresis ("ä") which in German should be sorted and looked at exactly as an "a" with diacritical character, but in Swedish has nothing to do with the character "a" except the look.⁵ While Sweden could implement one set of rules under .se and Germany another under .de, what rules should be used for generic TLDs such as .com? These questions could require serious negotiations between nation-states and their language experts to find compromises.

Domain names are intended to unambiguously associate a name to an IP address. This does not work effectively in the case of "homoglyphs" where domain names contain characters which are visually indistinguishable at quick inspection. Homoglyphs are present in domain names using LDH ASCII script - e.g. a lower case "l" and an upper case "I" look the same in some typefaces, so that a website URL like <http://paypai.com> could be written to be indistinguishable from <http://paypal.com> if cases are mixed in a browser address bar.⁶ However, the number of visually similar characters is greatly expanded when a large variety of scripts can be used.

To some extent, security issues will be mitigated by the fact that individual users will be able to distinguish areas of risk within their own scripts. However, the DNS is globally accessible and contains no reference to a user's context, so IDNs offer many more opportunities to be exploited in fraudulent ways. For example, characters from different scripts which are visually equivalent can be used to launch "phishing" attacks and mislead users into thinking an Internet site is genuine. Michael Everson notes that, "in Burmese the digit zero and letter wah are 100 percent identical in every font and there is no getting away from that."⁷

As the various cases were explored it became clear that rewriting the underpinnings of the DNS to account for every script and language would be impossible, and development of IDNs requires a number of compromises and balancing of priorities in the technical, cultural, and organizational arenas. The complexity of these issues has led the Internet Architecture Board (IAB) to determine that it would be extremely disruptive to transition the entire DNS to a

⁵ Signposts in Cyberspace: The Domain Name System and Internet Navigation – Committee on Internet Navigation and the Domain Name System: Technical Alternatives and Policy Implications, National Research Council, 2005. <http://www.nap.edu/catalog/11258.html>

⁶ Another example is a digital numeral "0" and upper case "o".

⁷ <http://www.icann.org/meetings/vancouver/captioning-idn-workshop-30nov05.htm>

fully IDN-compliant system, despite numerous limited implementations in non-ICANN namespaces. Instead, two testbeds are being developed by ICANN that are trialling the impact of global IDN implementation on the DNS and Internet usability.

ICANN's IDNA and DNAME testbeds

IDNA (RFC 3490) maps Unicode characters to ASCII-compatible encodings at the application level. Under the IDNA system, characters in a domain label are first normalized according to Unicode specifications through a function called "nameprep". Unicode allows strings to be represented in multiple forms. Nameprep consolidates these strings into a preferred form that can make comparisons and indexing simpler. It also eliminates different labels that have the same linguistic meaning, although it cannot eliminate alternate representations entirely. For example, the string "fi" can be represented either as the characters "f" and "i" (U+0066 U+0069) or by the ligature "fi" (U+FB01). Nameprep will treat these as equivalent.

Following nameprep, the normalized names are turned into an ASCII-Compatible Encoding (ACE) format, also known as "punycode". This creates a new domain name containing only 7-bit ASCII (LDH) characters that can be sent through the DNS. This is then converted back to Unicode by applications on the other side of the "wire". The prefix "xn--" is added to the ASCII encoding to indicate that the domain label should be treated as IDN encoded. For example, the hypothetical domain label <http://日本.co.jp>, typed into a web browser, would be converted to <http://xn--wgv71a.co.jp> in punycode. Similarly, if the user follows a link to <http://xn--wgv71a.co.jp>, it would appear in an IDNA-aware browser as <http://日本.co.jp>. It should be noted that end users should not have to manipulate domain names in an xn-- encoded format, however these may appear when applications are not IDNA-aware.

The benefit of the client-side approach of the IDNA standard is that it is compatible with the existing DNS. However, in order for IDNA to function universally, all software applications that interact with a domain name must be upgraded to implement the IDNA standard, including browsers, email applications, word processors, operating system tools, etc.

Another downside to IDNA comes from requiring conformance at the application layer rather than within the DNS infrastructure, which results in a lack of control over implementation. Therefore, even though the guidelines are specific about how applications should implement IDNs, applications can still be produced which implement IDNs in unusual ways or not at all. The required extensions to browser operations and syntax are not standardized and not consistent across all applications, meaning different users may receive different results depending on which tools they are using, and the rollout of a universally available set of tools will take years, if it happens at all.

A more significant problem is that IDNA is not yet capable of being used by most email clients, and email is one of the most important functions of the DNS. This

is especially important because email very often includes users' personal names, whose accurate representation will be very important to the average user. It also functions as an important aspect of corporate identity - the domain section of a user's work email address will usually be read as the marker for the employer of that user.

Even with acceptance of IDNA proposals, the question remains open as to how multilingual domains would be mapped into the DNS. For example, is a Japanese language version of .com (say, .会社, company) an entirely new domain space requiring its own registration procedures; or does it map onto the existing .com domain space? In the latter case, should a user typing in <http://動物.会社> go directly to the <http://animals.com> website?

This last example relates to the most controversial IDN implementation being tested by ICANN: DNAME.⁸ DNAME is a type of DNS record used to map or rename an entire sub-tree of the DNS name space to another domain. Under the DNAME scenario, alternative representations in different scripts would be mapped to existing ASCII TLDs. From the point of view of existing registries controlling one or more TLDs, it is an attractive proposal because it means that, for example, Verisign would be able to offer .com in a number of different languages (with significant revenue implications) rather than .会社 being offered by an entirely new and different registrar perhaps based in Japan.

Another proposal being tested (NS-record) is equivalent to how current DNS entries are currently made for any new TLD. For these, an internationalized label in punycode format (for .会社, xn--6oq404h) would be inserted in the root zone. This means that .会社 would effectively be equivalent to a new TLD which could be managed in a completely distinctive manner (and by a different company) than .com. This would have a much greater economic and political impact on the domain name industry and its management.

What other systems are available for multilingual domain names?

An axiom of the Internet is that because there is no control enforced by legislation (only agreement and recommendation), people will use a new system if it works for them. This is the case whether or not a solution is the most technically efficient for the network as a whole. So a number of organizations have deployed solutions for multilingual navigation services which serve DNS-like functions. It should be noted that not all of these systems are in direct competition with the ICANN DNS. Many have simply emerged to fill a market need and have expressed interest in migrating to global standards if/when they are developed. However, for some successful companies there is little incentive for them to migrate to standards such as IDNA which are less effective for their users.

⁸ This function was defined in RFC 2672 in 1999.

Providers of IDN systems include China Internet Network Information Centre (CNNIC) in China, i-DNS.net, a plugin-based architecture; Japan Network Information Centre (JPNIC), who have registered over 60,000 domain names in Japanese, and Korean Network Information Centre (KRNIC), who have registered over 50,000 domain names in Hangul. These are mostly in accordance with IDN guidelines apart from CNNIC's implementation. Verisign have also established a testbed for IDNs at the second level.

While most of the examples above relate to deployment of IDNs at the second level and beyond of the DNS, a number of countries and regions have begun testing deployment of IDN TLDs outside of the ICANN system. A number of technical experts, including the IAB have reiterated the importance of a single and authoritative root. From the IAB's perspective, "there is no getting away from the unique root of the public DNS." That sentiment was reiterated by a recent report from ICANN's Security and Stability Advisory Committee on alternate root systems.⁹ However, former ICANN board member Karl Auerbach claims that the ICANN report "does not raise any technical reason why as a technical matter there can not safely coexist on the net several different DNS naming spaces - which may or may not be consistent with one another - each dangling from a different DNS root."¹⁰ Ironically, the telephone system numbering plan works in such a distributed fashion with no single technical root. There is only logical coordination by the International Telecommunication Union (ITU) - with, depending on the context, national regulation or market forces taking care of the rest. This distributed policy model has made innovation possible in national and regional contexts (e.g. national freephone numbers) while still supporting more general global resolution.

Different DNS name spaces can co-exist, provided there is coordination and cooperation amongst all namespace owners to avoid collision and to mutually cross-resolve each other's namespace to their respective end users to preserve universality. This is what a number of Arabic, Chinese, Farsi, Hebrew, Korean, Russian IDN TLD operators are currently trying to do under the coordination of the MINC, to set up a universal resolution system that will take their fragmented namespaces and enable cross-resolution.

What are the differences of opinion on IDN implementation?

The biggest issues with IDN are not technical, but political - they relate to the relative priority given to different users and their needs. This is for two reasons:

Firstly, the DNS is not designed for computers but for people. The Internet would be simpler from a technical perspective if the DNS did not exist and everyone used IP addresses to identify the computers they wished to reach, much as we use telephone numbers today. Therefore, the discussion about how the DNS should

⁹ <http://icann.org/committees/security/alt-tlds-roots-report-31mar06.pdf>

¹⁰ <http://www.cavebear.com/cbblog-archives/000245.html>

function is primarily about humans and how we interface with the technology. In particular, there are competing ideas of “the user” and what is logical and effective for the user which are at stake in the discussion. The average user themselves has little understanding of the DNS (especially compared to the “average user” in 1985, who was often a member of the technical community) so it is left to others to advocate on their behalf in technical and policy frameworks.

Secondly, the DNS as currently controlled by ICANN and the Internet Assigned Numbers Authority (IANA) is not the only mechanism by which users navigate the Internet, though it is by far the most prominent. As well as the various alternate-script systems mentioned above, many important software applications such as Skype and AOL Instant Messenger have navigation and identification systems that do not require the DNS, but instead offer users their own private directories to facilitate communication and file transfer. Similarly, search engines have to some degree reduced dependence on the DNS to identify organizations and individuals on the Internet.

Therefore, the single most important question is not whether alternate systems are possible, but whether the overall benefits of a single namespace and unique root outweigh the benefits of allowing users to use the language of their choice in the DNS.

In general, there are two identifiable groups of opinion in this discussion. They are not completely polarized, and individuals and organizations may tend more toward one or the other position.

The unique root

The first and most powerful bloc consists of ICANN and the Internet Engineering Task Force (IETF) and its most influential stakeholders: registry owners, the US/European private sector, engineers in this private sector and civil society, and the US Government - those for whom the current system works well, or at least better than practical alternatives. For slightly different but compatible reasons, they prioritize the stability of the existing system and the need for a unique root. For these communities, the introduction of IDNs must proceed slowly and without disrupting the existing system.

- For the registry owners, a unique root maximizes the value of their investment in both the existing namespace and in their participation in ICANN and technical bodies such as the IETF. Multiple roots will act as competition.
- For the private sector, a single and unique root makes it significantly easier to manage intellectual property considerations around domain names and navigation systems. Transnational companies resist new TLDs as they fear they have to register their names in the new TLDs whenever they emerge in order to prevent passing off by others. When they neglect to register, others will capitalize on their omission and cybersquat or pass off, or generate fake sites for phishing and other nefarious activities, thereby confusing end users.

However, it has been noted that in the case of IDN TLDs, this will generally be taken care of by their own branches in the countries which they operate. Other than additional expense, it is within the standard operating procedure of such companies to detect for passing offs, and well within existing intellectual property laws to handle such cases expeditiously.

- For engineers involved in the DNS, it is self-evident that the DNS only works where there is a single and unique root, because that is how the system was designed. The implications of deploying multiple public DNS roots would, in the words of RFC 2826, “raise a very strong possibility that users of different ISPs who click on the same link on a web page could end up at different destinations, against the will of the web page designers.”¹¹ This would increase the chances of fraud and could reduce the overall usability and coherence of the system.
- The US Government's National Telecommunications and Information Administration (NTIA) noted in 2005 that “the United States is committed to taking no action that would have the potential to adversely impact the effective and efficient operation of the DNS and will therefore maintain its historic role in authorizing changes or modifications to the authoritative root zone file.”¹² Although many members of the technical community would prefer to see this oversight fully delegated to ICANN, it nevertheless means that there is a shared interest with the US Government in maintaining a single, authoritative root.

Prioritizing multilingualism

The second body of opinion sees the lack of progress on IDNs as a critical issue that prevents the non-English speaking majority of the world from making effective use of the Internet. In this group are the IDN providers, technical bodies working on script encoding (especially from Asia), governments whose official languages do not use Roman script, cultural rights advocates, and many members of civil society in North America and Europe. For these communities, the priority is to implement systems that allow users to navigate using a range of scripts, and that the decisions about what is possible should be made by the language communities themselves. They point out that international business still takes place without a single shared language, and that different language communities can negotiate interfaces between their relatively discreet cultural and economic systems.

- IDN providers want to gain access to a valuable market for domain registry services that has so far been located in the US and Europe.
- Various technical bodies in Asia who have navigation systems in particular scripts would like

¹¹ <http://www.faqs.org/rfcs/rfc2826.html>

¹² http://www.ntia.doc.gov/ntiahome/domainname/USDNSprinciples_06302005.htm

to see this work taken up on the Internet as a whole. They are impatient with delays and a perceived lack of commitment from ICANN/IETF to this issue.

- Governments whose official languages do not use Roman script are uncomfortable with decisions on domain name deployment sitting with a private US-dominated body. They are used to having control over things like language rules and mandating how languages and scripts should be used.
- Cultural rights supporters (primarily in civil society) see the use of one's own language as a basic human right, which is not outweighed by ease-of-use or security considerations proposed by predominantly English-speaking groups for whom the current system is viable.

Conclusion: Multilingual domains - Future scenarios

Acknowledging this tension, the issue of global interoperability within the DNS' limited capabilities must be balanced against the detrimental effect of many users not being able to effectively use their own language in the construction of information exchange systems. The perspectives voiced in this debate are largely determined by how easily one is able to use the English language. For many users, the benefits of the "globally interoperative" DNS are theoretical rather than practical. Furthermore, because the use of the ICANN-controlled DNS is by recommendation rather than law, it is always possible for new systems to be deployed that work for particular script communities. If they work, people will use them, regardless of opinions about their suitability for the Internet as a whole.

Comments in 2004 by former IAB Chair John Klensin have acknowledged that the approach taken by IDNA, while better than any other actually existing alternatives in his view, suffers from significant limitations. If IDNs are this hard and do not solve the problem... maybe it is time to go back to the problem and do some serious thinking about models which would be "non-DNS" or "above-DNS".¹³

The maintenance of a single unique namespace for a truly multilingual global medium brings unprecedented challenges. The attempt to fully internationalize the DNS would be more complex than the massive project undertaken by Unicode, finding a common technical encoding for scripts (itself far from uncontroversial). A single, multilingual namespace would also be the first attempt to mediate between competing uses of the same words on a global scale, some of which had previously been allowed to exist in different languages. A single system which serves a range of language groups would require mechanisms to effectively negotiate between different priorities of language scholars, trademark owners, standards bodies, and technical engineers, to name a few. In particular, the role of nation-states in formalizing language has never

been as complex as a single unique namespace would require.

To effectively support this negotiation would require more resources than existing bodies such as ICANN currently have available for the task. It is only in 2005 that ICANN committed significant resources to the problem with a President's Advisory Committee working on the subject and the proposal for IDN testbeds to be put in place. For those who have been working on alternative systems, there is little confidence in the ability of the established regime to achieve their interests. If the bodies in control of the existing DNS wish it to remain the default Internet navigation standard, they must not only work to implement IDNs in areas where it is economically lucrative, but provide an effective participation mechanism for all the stakeholders in this complex domain.

Danny Butt is a partner at Suma Media Consulting <<http://www.sumamedia.com>>

Additional Reading

APDIP e-Note 1 - Voices from Asia-Pacific: Internet governance and sustainable human development – Akash Kapur and Christine Apikul, 2005.
<http://www.apdip.net/apdipenote/1.pdf>

Internet Governance: A Primer – Akash Kapur, 2005.
<http://www.apdip.net/publications/iespprimers/eprimer-igov.pdf>

Internet Governance: Asia-Pacific Perspectives – Danny Butt (ed.), 2005.
<http://www.apdip.net/publications/ict4d/igovperspectives.pdf>

Examining Top Level IDNs – John C Klensin, 2005.
<http://www.icann.nl/announcements/examining-top-level-idns-17nov05.pdf>

ITU and UNESCO Global Symposium on Promoting the Multilingual Internet, Geneva, 9-11 May 2006.
<http://www.itu.int/ITU-T/worksem/multilingual/>

Multilingual Internet Names Consortium (MINC)
<http://www.minc.org>

Signposts in Cyberspace: The Domain Name System and Internet Navigation – Committee on Internet Navigation and the Domain Name System: Technical Alternatives and Policy Implications, National Research Council, 2005.
<http://www.nap.edu/catalog/11258.html>

Suggested Practices for Registration of IDNs – The Internet Society, 2005.
<http://www.ietf.org/rfc/rfc4290.txt>

¹³ <http://ws.edu.isoc.org/workshops/2004/ICANN-KL/ICANN-ISOC-KL-IDN-part2.ppt>