

Internationalised Domain Names and Nominet Consultation Document

Clive Feather

Draft 1, 2004-12-06

Consultation Summary

Domain names are currently limited to the letters A-Z, the 10 digits, and hyphens. This excludes accented letters and other character sets. A new approach called IDNA extends names to allow all the writing systems used around the world.

The IDN Working Group has been looking at options for going forward with IDNA, ranging from doing nothing to providing full support. A number of secondary matters, from support facilities to dispute resolution, have also been looked at.

This consultation is being held to solicit much wider input from the stakeholder community on these matters.

Introduction

The “traditional” Internet was mostly designed by English speakers and therefore was based around the ASCII character set. When this was applied to domain names, this meant they were limited to the 26 “usual” unaccented Latin letters, the 10 “Arabic” digits, and hyphen (plus, of course, dot between components). In addition, it was decided that the case of letters would not be significant – **example**, **EXAMPLE**, and **eXaMPIE** are all equivalent.

Now that the Internet is truly international, it is unreasonable for domain names to be limited to these 37 characters (referred to in the rest of this paper as the “LDH” – letters, digits, and hyphen – character set). Even in other countries that use the Latin alphabet, people want to be able to spell their names properly, meaning that it is necessary to cope with accents like â and additional letters like ß. When we look further afield, there are many places where either non-Latin alphabets are used – such as Greek, Cyrillic, or Arabic – or other forms of writing – such as Chinese or Thai. Even within the UK, many citizens use writing systems other than the Latin alphabet; even if they are fluent with the latter, they still use other scripts in everyday life.

It is unreasonable to expect all these people to use a completely different writing system for domain names than for their normal everyday life, and the Domain Name System needs to adapt.

Note that it would be naïve to dismiss this as a “minority” or “foreign” problem – neither English nor Welsh can be properly written using only the LDH characters.

The IDNA approach

The technical approach chosen by IETF to solve this problem is called *Internationalizing Domain Names in Applications*, or *IDNA*; it is defined in RFC3490. In this approach, the DNS infrastructure is left completely untouched. Instead, individual applications are upgraded to recognise domain names in Unicode¹. When they are given such a domain name (an *Internationalized Domain Name*, or *IDN*)², they convert all labels (the components of domain names separated by dots) that are not written in the LDH set into *ACE* (*ASCII Compatible Encoding*), which is an encoding of Unicode using only the LDH characters. The resulting “IDN-unaware” domain name is then passed to the DNS in exactly the same way as normal domain names. An ACE label can be easily identified because it begins with **XN**— (that is, X N followed by two hyphens). For example:

User enters:	www.nestlé.co.uk	www.êpô.org.uk
Application converts to:	www.xn--nestl-fsa.co.uk	www.xn--bdau3a.org.uk
Nominet registry entry:	xn--nestl-fsa (in co.uk)	xn--bdau3a (in org.uk)

Limits on this consultation

This consultation does not envisage Nominet defining a competing approach to IDNA. Whether or not it is perfect, IDNA is likely to be the solution adopted by much of the world and there is no good reason to do something different.

¹ Otherwise known as ISO10646. See (<http://www.unicode.org>) for a full description.

² Not every possible Unicode string is a legitimate IDN, just as not every string made up of LDH characters is a legitimate domain name at present – for example, names beginning with hyphens are not legitimate. In particular, not every Unicode character is permitted in IDNs.

This consultation does not discuss the issue of multi-lingual support at Nominet (that is, support in French or Urdu, as opposed to support in English for names using French or Urdu characters). While it may appear that there are links between the two topics, they are weak enough that they should be addressed separately; this consultation looks only at domain names.

This consultation discusses 3rd level domain names because they form the vast majority of IDNA registrations addressed by Nominet will be of this form. Should Nominet register names at deeper levels, the same rules will apply to them. Second level domains (SLDs) are a different matter. If Nominet registers IDNs, there is no technical reason to forbid SLDs from being ACE labels. However, the policy issues are likely to be much more complex. Since there are separate processes for registration of SLDs, this consultation does not examine them further.

Issues for Nominet

Nominet's rules currently forbid the registration of names where the third and fourth characters are both hyphens, and therefore forbid the registration of ACE labels or strings that look like them. The IDNA approach means that the core DNS is not affected by IDNA and, therefore, there are no key technical issues with allowing registration of ACE labels; it is merely a case of removing the current restriction. However, the issues facing Nominet are far wider than the technical ones. Rather, they revolve around questions of customer support, dispute resolution, and similar matters.

This consultation offers three basic options:

1. Status quo: ACE labels cannot be registered.
2. Minimum: ACE labels can be registered, but Nominet does not discuss IDNs.
3. Full: Nominet provides a range of facilities for handling both IDNs and ACE labels.

A number of detailed variations are also offered.

Option 1 – Do Nothing

Under option 1, Nominet does nothing and continues to forbid the registration of ACE labels.

The advantage to this option is that no new resources need to be expended, at least in the short term.

The major disadvantage is the same as with any conservative decision: if the rest of the world moves to IDNA, Nominet will be left behind. Over time it will be less attractive to potential registrants, there will be more queries about why the facility is unavailable, and pressure to reverse this decision will increase.

There may also be a legal issue related to discrimination? Is there a requirement to support at least Welsh and/or various EU languages? Or scripts used by a significant number of people in the UK? Even if Nominet itself isn't liable, are some of our customers (particularly local authorities)?

Option 2 – ACE registrations but no new infrastructure

Under option 2, Nominet allows the registration of ACE labels but provides no other facilities, treating them like any other domain name. Any form of specialised support, such as a WHOIS that understands IDNA, would be left to tagholders.

The advantage to this option is that it provides support for IDNA at minimal effort to Nominet. The ability of tagholders to provide specialised support can also be seen as an advantage.

The disadvantage is that Nominet can expect to receive queries and complaints involving IDNA names. In particular, registrants may not understand the ACE label concept well and are likely to confuse names differing only in the use of accents, while staff may have difficulty with descriptions of non-Latin characters, particularly on the telephone. There is also a risk of increased numbers of DRS disputes related to these same issues. Failure to provide proper support for registrants of these names can be seen as providing them a second-class service.

Option 3 – Full support for IDNA

Under option 3, Nominet not only allows the registration of ACE labels but also provides a range of support for them. For example, certificates could carry both the IDNA and ACE name, the WHOIS could provide a range of ways to enter IDNA names, and staff would be trained to deal with the various issues involved.

Under this option Nominet would be fully embracing the opportunities associated with IDNA. This should benefit all stakeholders: registrants because of the increased choice without loss of convenience, tagholders because of the potential new markets and the opportunities it provides, and Nominet because it will be seen as continuing to provide a world-leading registry.

The disadvantages of this option are that Nominet will have to spend significant amounts on IDNA support, and that tagholders may have to be prepared to handle IDNA (more than just accepting ACE labels) in order to retain their position in the market.

Finally, there is the risk that IDNA will not obtain the “critical mass” required to justify the efforts put into it, meaning that the money and effort put into supporting it will have been wasted.

Consultation question 1: Do you think these three options (with the variations outlined below) encompass the possibilities?

Consultation question 2: Which of the three options do you think Nominet should adopt, and why?

Consultation question 3: Are you aware of any legal issues that would affect Nominet’s handling of IDNA?

Consultation question 4: Do you have any other comments on the options and their advantages and disadvantages?

Supported character sets

IDNA is not a free-for-all: it still puts some restrictions on the characters that can appear in domain names. For example, spaces cannot appear, and where two different labels have the same meaning, only one will be acceptable (for example, German “ß” cannot appear; “ss” must be used instead).

Nominet has the choice of allowing all possible labels that IDNA allows, or of restricting the set in some way³. For example:

- Names could be listed to those characters described as “alphanumeric” by Unicode. This would forbid domain names like £.co.uk .
- Diacritical marks (accents and similar) could be forbidden. Thus names like nestlé.co.uk and châteaux.org.uk would be prevented, as being too similar to nestle.co.uk and chateau.co.uk . Alternatively names with accents would be permitted, but each would block registration of any other name differing only in the accents used (or not).
- Characters could be limited to those in certain language blocks (e.g. Greek and Urdu could be permitted while Cyrillic and Thai are forbidden).

Consultation question 5: If Nominet adopts IDNA, should it restrict the characters allowed in domain names over and above the limitations of IDNA itself?

Consultation question 6: If so, what restrictions should be applied, and why?

Additional facilities

There are a range of facilities that Nominet could provide when supporting IDNA. For example:

- The automaton, which is text-based, could allow various ways of representing IDNs rather than being restricted to ACE labels. For example, “ \hat{g} ” could be entered as “[g^]”, as “[011D]” (its Unicode encoding) or even as “[g circumflex]”.
- Nominet could provide web tools for converting between IDNs and ACE labels (e.g a form where someone types in an ACE label and the corresponding IDN, if any, is displayed).
- The WHOIS web server could allow input of IDNs in various ways.
- Certificates could include both forms of the name (and an explanatory text). This would clarify the status of ACE labels.

Consultation question 7: What other support for IDNs can you suggest?

Consultation question 8: For each of these facilities, should they be provided regardless of cost, not be provided at all, or be provided only if cheap and simple to do?

Sunrise period

The opening up of a new range of domain names can be expected to produce a burst of demand. This could lead to overloading of Nominet’s systems, and to additional disputes as stakeholders race to register names that appear similar to existing ones, either because they failed to obtain them in the past, to protect their own brand, or even in attempted cybersquatting. Clearly the “sunrise” period needs to be managed in some way.

Two possible methods for doing this are:

- Provide a pre-registration period where existing registrants can apply for names “similar” (defined in some manner to be determined) to their own.

³ This consultation is not concerned with the technical details of how the limitation is set; an unambiguous definitive rule will be chosen once a policy decision has been made.

- As with me.uk, set an initial high price which then reduces on a sliding scale over a few months to the normal level.

Both could be used, of course.

Consultation question 9: What approach (not limited to the two above) do you think should be used for managing the sunrise period, and why?

Dispute resolution

IDNs will, of course, be subject to the same dispute resolution issues as other names. However, two additional problems have been noted:

- How can an Expert determine whether two names – in a script or language not familiar to him – are similar enough to introduce a risk of confusion?
- Could a name in some script (e.g. Arabic, with its flowing curves) look like a logo or other trade mark, and does this introduce problems?

We would particularly like comments from DRS Experts on how IDNA could affect their work.

Consultation question 10: Do you think either of these are a significant problem?

Consultation question 11: Are there any other ways in which IDNA will affect dispute resolution and the DRS?

Other matters

The Working Group does not claim to have a monopoly of expertise in this matter. We realize that we may well have overlooked other possibilities or matters for concern.

Consultation question 12: Are there any other issues relating to IDNA that should be taken into account, or that you want to mention?

Next steps

Following this consultation, the IDN Working Group plans to produce a policy recommendation to be presented, via the Policy Advisory Board, to the Nominet Board. It will then be up to the latter to decide how to proceed.

How to respond

[[Insert boilerplate]]