

PAB Consultation Paper: Internationalised Domain Names and Nominet

1. Overview

Nominet's Policy Advisory Board (PAB) is a group of representatives of the UK Internet community, including government, industry and others which advises Nominet's board on the policy direction the registry should take. The PAB has set up a working group on Internationalised Domain Names and is now seeking your comments and views.

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 has the potential to allow "International Domain Names" which contain any of the writing systems used around the world. This could allow domain names with accents (<café.co.uk>), or with letters from other alphabets (<êpô.org.uk>) or even totally different writing systems, such as Chinese.

The PAB is currently considering whether and how IDNA could be introduced for the .uk domain name.

This document necessarily refers to a number of technical matters. However, it is not necessary to understand the technical issues to be able to respond to this consultation, and Nominet welcomes responses from anyone interested in the topic – particularly likely end-users of Internationalised Domain Names – whatever their level of technical knowledge.

The consultation deals with three topics.

- The first is the degree to which Nominet should support Internationalised Domain Names (from 'Not at all' to 'Totally').
- The second is which letters and characters should be allowed (e.g. just allow Welsh, or allow everything).
- The third deals with related matters such as the effect on the Dispute Resolution Service and other systems.

1.1 Note

[0] Paragraph numbers, in brackets, are provided for the convenience of respondents wishing to refer to a particular point, and have no other significance. A glossary of terms is included at the end of this document.

2. Summary of Questions

There are a number of specific questions set out in the consultation. They are reproduced together here.

Question 1: Do you think these four primary options [in section 7, below] (with the variations outlined below) encompass the possibilities?

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

Question 3: Are you aware of, or do you have experience of, any legal issues that could affect Nominet's policy choice as regards IDNA?

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

Question 5: If Nominet adopts IDNA, should it restrict the domain names that can be registered (either in terms of the characters allowed in domain names, or otherwise) over and above the limitations of IDNA itself?

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

Question 7: Are there other forms of support for Internationalised Domain Names that you would like to see Nominet provide?

Question 8: For each of the IDN support facilities mentioned in this section, and also any you suggest in answer to question 7, should they be provided regardless of cost, not be provided at all, or be provided only if cheap and simple to do?

Question 9: How should the sunrise period operate, and why?

Question 10: Do you think any of these dispute resolution issues are a significant problem?

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

Question 12: Would you expect to use IDNs as your primary domain, or would you only be expecting to register them for “defensive” reasons?

Question 13: Are there any other issues relating to IDNA that should be taken into account, or that you want to mention?

3. Introduction

[1] The “traditional” Internet was mostly designed by English speakers and therefore was based around what might be called the ‘simple English latin’ alphabet of the 26 “usual” unaccented Latin letters, the 10 “Arabic” digits (0-9), and hyphen (plus, of course, the dots). In addition, it was decided that capitalisation of letters would not be significant, so that **internet**, **INTERNET** and **iNtErNeT** are all equivalent.

[2] Now that the user base of the Internet is truly international, it is widely believed that there is a need for domain names to go beyond these 37 characters (referred to in the rest of this paper as the “37 core symbols”). Even other Western European languages make use of characters other than the 37 core symbols (e.g. with accents such as â, or additional letters like ß). Looking further afield, non-Latin alphabets are used – such as Greek, Cyrillic, or Arabic – or other forms of writing – such as Chinese or Thai. The nature of the UK is such that many citizens speak and write in these non-English languages in their everyday life.

[3] Strictly, it would be naïve to dismiss this as a “minority” or “foreign” problem – as there are words in both English and Welsh that require letters beyond the 37 core symbols.

[4] Allowing domain names with characters from these other alphabets or writing systems has the potential to allow a huge range of domain names relevant to the community that uses them.

[5] On the other hand, allowing thousands of different symbols in domain names can lead to confusion, and there are practical difficulties which Nominet would face if someone had a domain name written in e.g. Japanese.

4. The IDNA approach

[6] The technical approach chosen by the international technical community to solve this problem is called *Internationalizing Domain Names in Applications*, or *IDNA* (which technical readers can find in IETF RFC3490, but which is summarised here). In this approach, the existing Domain Name System infrastructure is left completely untouched. Instead, individual applications (e.g. web browsers such as Internet Explorer or Mozilla Firefox) are upgraded to recognise domain names written in (almost) the full range of known symbols. When a user types in a domain name which uses characters which are not in the 37 core symbols (an *Internationalized Domain Name*, or *IDN* such as <êpô.org.uk>), the application will convert this IDN into a domain name which only uses the 37 core symbols (in that case <xn--bdau3a.org.uk>). The way in which the application does this conversion is set out in rules which the international technical community have devised.

[7] For our purposes, all that matters is that what comes out at the other end is, in effect, a “plain” domain name which may look like gibberish to the human eye, but obeys all the existing rules for the domain name system, and only contains characters from the 37 core symbols. Because of that, the current Domain Name System can deal with it in the usual way.

[8] These “plain” domains are properly known as “ACE labels” (ASCII Compatible Encoding). They can be easily identified because they begin with **XN--** (that is, X N followed by two hyphens). For example:

Example 1:

User enters an IDN: **www.café.co.uk**
 Application converts to an ACE label: **www.xn--caf-fsa.co.uk**
 Nominet registry entry: **xn--caf-fsa.co.uk**

Example 2:

User enters an IDN: **www.êpô.org.uk**
 Application converts to an ACE label: **www.xn--bdau3a.org.uk**
 Nominet registry entry: **xn--bdau3a.org.uk**

There is therefore a difference between an IDN (which has characters other than the 37 core symbols) and an ACE label (which does not). However, for most of the purposes in this consultation paper, we will just use the term ‘IDNA’ to convey the idea of the internationalised domain name system together. Where we refer to just one or the other, there will be a reason for doing so.

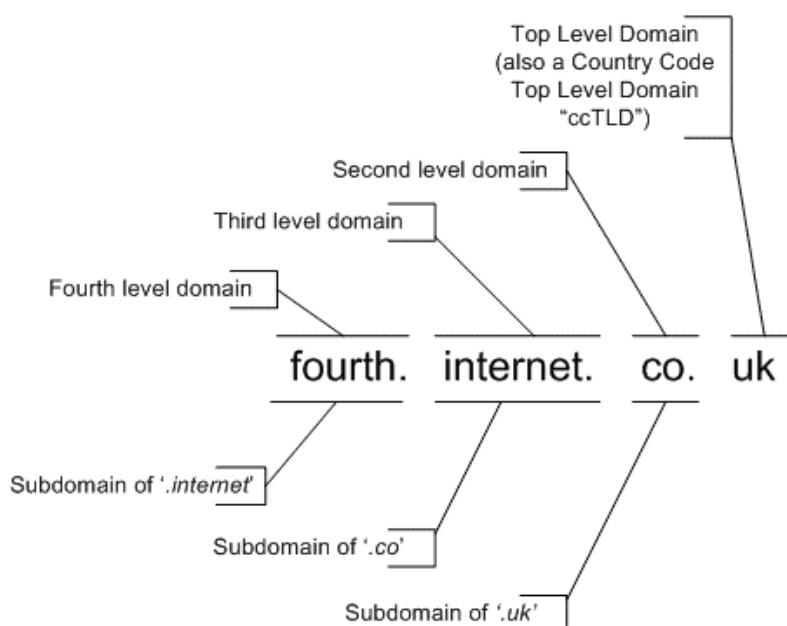
5. Limits on this consultation

[9] 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. This paper therefore does not consider whether an alternative system would be better.

[10] This consultation deals with whether or not Nominet should allow IDNA, but not the more general issue of how to support speakers of languages other than English e.g. whether Nominet (would have to employ large numbers of staff who speak those languages).

[11] This consultation discusses 3rd level domain names (Figure 1 explains the levels of a domain name), because they form the vast majority of IDNA registrations that will be addressed by Nominet, but the same rules will apply to fourth level domains etc. if Nominet records them.

Figure 1



[12] If Nominet registers IDNs, there is no technical reason that second level domains could not also be in non-English formats but the other issues are likely to be more complex and this consultation does not deal with them.

6. Issues for Nominet

[13] Since all IDNAs start with "xn--" Nominet's rules currently prevent any domain names in this format being registered, pending the outcome of the international technical community's development of IDNA. In basic technical terms, all that is required in order to allow IDNA would be to remove this block. In practice, this will not be done until the serious issues that IDNA pose are considered.

7. The Options

[14] This consultation offers four basic options:

1. Do nothing.
2. ACE registrations but no new infrastructure.
3. Automated support for IDNA.
4. Full support for IDNA.

A number of detailed variations are also offered.

7.1 Option 1 – Do Nothing

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

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

[17] 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 Nominet will become less attractive to potential registrants as we will not be providing what our customers can get elsewhere.

[18] By doing nothing we could effectively be creating a barrier to enabling our customers to serve their communities to the full, particularly if they have an obligation to provide access in a non-english language e.g. if they have an obligation to support Welsh, or other EU languages, or scripts used by a significant number of people in the UK? While legal advice may be needed, an indication of areas that might have been overlooked, and of existing experience of stakeholders, will help inform the process.

7.2 Option 2 – ACE registrations but no new infrastructure

[19] Under option 2, Nominet allows the registration of ACE labels but provides no other facilities, treating them like any other domain name. Nominet would, in effect, deal with <xn--caf-fsa.co.uk> exactly the same as it does with <cafe.co.uk> and it would be up to the registrant involved and/or their tagholder to understand what <xn--caf-fsa.co.uk> means and what it would 'turn into' when used in the IDNA system. Nominet systems, such as the WHOIS, would not return a result for <café.co.uk> even if <xn--caf-fsa.co.uk> was registered: all this sort of support would be left to tagholders.

[20] 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.

[21] 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 the probability 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.

7.3 Option 3 – Automated support for IDNA

[22] Under option 3, Nominet not only allows the registration of ACE labels but also provides a number of simple automated support facilities for them. For example, letters to registrants could carry both the IDN and ACE name, though with a note saying that Nominet will need to know the ACE label for all communications. Equally, a 'translation' tool could be provided that would show what the ACE label equivalent of an IDN was and the WHOIS could provide a range of ways to enter IDN names. However, staff would continue to only work in terms of the ACE label i.e. if a telephone caller told the Nominet staff member that they were the registrant of <êpô.org.uk>, the Nominet staff member would be unable to deal with them until the caller told them that the ACE label version of this name is <xn--bdau3a.org.uk>. Quite how the registrant would know that is part of the aim of this consultation.

[23] Under this option Nominet would be dealing with some of the problems of option 2 without invoking all the disadvantages of option 4. Conversely, it retains some of the problems of the former while abandoning some of the benefits of the latter. As with any compromise, the question is one of finding the optimal point.

7.4 Option 4 – Full support for IDNA

[24] Under option 4, Nominet not only allows the registration of ACE labels but also provides a full range of support facilities for them. As well as letters carrying both the IDNA and ACE name, for example, staff would be trained to deal with the various issues involved and would be ready to deal with registrants on a fully-IDN basis.

[25] Under this option Nominet would be seen to be fully embracing the opportunities associated with IDNA. This could 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.

[26] 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.

[27] 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 four primary options (with the variations outlined below) encompass the possibilities?

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

Consultation question 3: Are you aware of, or do you have experience of, any legal issues that could affect Nominet's policy choice as regards IDNA?

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

7.5 Supported character sets

[28] 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 because the "ß" character is seen as simply a different way of writing "ss". However, where two characters look the same but have either different origins (e.g. Latin "A" and Cyrillic "А") or have different meanings (e.g. Greek capital rho and Cherokee tlv, both of which look like the English "P"), both are allowed to appear in IDNs.

[29] Nominet has the choice of allowing all possible labels that IDNA allows, or of restricting the set in some way¹. For example:

¹ 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.

- Names could be limited to alphanumeric symbols (there are international standards which define what ‘alphanumeric’ means in this case). This would forbid domain names like < ℥.co.uk>.
- Marks with 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). Obviously this would limit the number of IDNA issued, since the number of unaccented words registered far exceeds the number of accented words currently registered (i.e. none, because they are forbidden).
- Characters could be limited to those in certain language blocks (e.g. Greek and Urdu could be permitted while Cyrillic and Thai are forbidden) limiting the problem highlighted in paragraph [28]. However, restricting .uk IDNA to certain language blocks could be contentious if the restrictions seem arbitrary. Some possible restrictions would include: limiting it to official languages in the UK (i.e. characters needed for Welsh, Scots Gaelic etc.), limiting it to characters needed for the official languages of the EU, or limiting it to character sets used by more than 1% of the UK population. Quite how that last test would be validated (from census records?) is not clear.

Consultation question 5: If Nominet adopts IDNA, should it restrict the domain names that can be registered (either in terms of the characters allowed in domain names, or otherwise) over and above the limitations of IDNA itself?

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

7.6 Additional facilities

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

- 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. It could also display both IDN and ACE forms of ACE labels.
- Letters confirming registration could include both forms of the name (and an explanatory text). This would clarify the status of ACE labels.

Consultation question 7: Are there other forms of support for IDNs that you would like to see Nominet provide?

Consultation question 8: For each of the IDN support facilities mentioned in this section, and also any you suggest in answer to question 7, should they be provided regardless of cost, not be provided at all, or be provided only if cheap and simple to do?

7.7 Sunrise period

[31] 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.

[32] Four possible methods (which are not mutually exclusive) 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.
- As with me.uk, set an initially high price which then reduces on a sliding scale over a few months to the normal level.
- No sunrise period (i.e. put in place no special provisions at all). This has been done successfully in Germany and Austria.

- Limit the number of names that can be applied for in a single application.

Consultation question 9: How should the sunrise period operate, and why?

7.8 Dispute resolution

[33] IDNs will be subject to the same dispute resolution issues as other names. However, three 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?
- If translation evidence is provided, how does an Expert decide between two, differing, opinions by translators as to the meaning or spelling of a word?

Consultation question 10: Do you think any of these dispute resolution issues are a significant problem?

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

7.9 Other matters

[34] Having an idea of expected use of IDNs would help formulate policy in this matter.

[35] One concern that has been expressed is that IDNA assists in “typosquatting” or cybersquatting e.g. that a company called “Internet Ltd” which currently has the <internet.co.uk> domain name might have to register <internet.co.uk>, <internet.co.uk>, <internet.co.uk> and <internet.co.uk> to protect itself from cybersquatters, because most users would not notice the change in the initial letter. Indeed, there are several letters that look identical, but are permitted as they are actually different letters from different languages (e.g. see para [28]).

[36] The other thing to remember about IDNA is that whether or not a registry supports IDNA is only part of the overall question. If you have a standard UK-issue keyboard, you will not have a key for “ö” although this is common in Germany, making an IDNA with these characters difficult to type. Japanese would be even harder. Equally not all software actually recognises IDNA at the moment: Internet Explorer v6 (the version current at the time of this paper) does not, while the second most common browser, Mozilla Firefox did offer IDNA support but has disabled it because of security concerns. For the average user, these problems might be insurmountable.

Consultation question 12: Would you expect to use IDNs as your primary domain, or would you only be expecting to register them for “defensive” reasons?

Consultation question 13: Are there any other issues relating to IDNA that should be taken into account, or that you want to mention?

8. How to respond

[37] In order to give input, please send an email to ids-consultation@nominet.org.uk, with the subject line ‘IDN Consultation’. All consultation responses will be published on Nominet’s website at <http://www.nominet.org.uk/Pab/PabConsultationPapers/IdnConsultation/IdnConsultationResponses.html>.

[38] We reserve the right to remove any materials which in our reasonable opinion are defamatory, offensive or unintelligible.

9. Next steps

[39] Following this consultation, the PAB’s 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.

[40] The process from now on is:

1. Consultation runs for three months (i.e. to 6 September 2005).
2. Consultation responses are summarised by the IDN Working Group who will then formulate a policy proposal and pass it to the PAB
3. The PAB will either endorse the proposal and forward it to Nominet's Board, or – if they have concerns – will ask the IDN Working Group to amend and resubmit it.
4. Nominet's Board considers the PAB's advice
5. Nominet considers how, in operational terms, the PAB policy recommendation can be put into effect.
6. Nominet takes these steps, which may require further PAB input if, for example, Nominet needs to change its Rules for domain registration to allow IDNs

10. Glossary

ACE	ASCII Compatible Encoding – a way of encoding IDNs into the 37 core symbols.
ASCII	American Standard Code for Information Interchange – the traditional character set used by most computers. It only contains 94 printable characters, not including any with accents or non-Latin letters.
DNS	Domain Name System – the basic system for managing domain names.
DRS	Nominet's Dispute Resolution Service.
IDN	Internationalized Domain Name.
IDNA	Internationalizing Domain Names in Applications, a process defined in RFC3490.
IETF	Internet Engineering Task Force – the body which sets the technical standards for the Internet.
Latin	In the context of character sets, characters based on the Latin alphabet. It includes the 26 letters of the modern English alphabet, similar letters with accents, and additional letters used in some western European languages (e.g. ð), but not Greek, Cyrillic, Chinese, or other character sets not derived from Latin.
RFC	Request For Comments – despite the name, a series of documents which includes within it most of the formal technical standards for the Internet.
SLD	Second Level Domain – e.g. co.uk or ltd.uk .
Tagholder	A person or company who manages a .uk domain name on behalf of an end-user and who can use Nominet's automatic registration system (the Automaton) – all Tagholders have a contract with Nominet..
Unicode	Otherwise known as ISO10646 (see http://www.unicode.org) – a character set for computers that contains every character in common use throughout the world.
WHOIS	The automated system that can be used to determine the registrant of any domain name issued by Nominet.