

IDN
technical specifications

- Version 1.0 -
February 3rd, 2012

afnic

Table of content

1. Foreword	3
1.1. Reference documents	3
1.2. Decision of the Board.....	3
1.3. Terms subject to prior review	3
1.4. Brief backgrounder on IDN technology	3
1.5. Warning	4
1.6. Terms and definitions.....	5
2. Table of accepted characters	6
3. Use of Unicode versions vs. LDH versions	8
4. Grandfathering Rule	8
4.1. Example 1	8
4.2. Example 2	9
5. EPP specifications	9
5.1. Example of an IDN query using EPP	10
5.2. Example of response.....	10
6. Web Interface Specifications.....	11
7. DAS specifications	11
7.1. Example of an IDN query in its Unicode form	11
7.2. Example of an IDN query in its ACE form.....	12
7.3. Example of an IDN query in its Unicode form during the sunrise period	13
8. Whois port 43 Specifications	14
9. Notification e-mail specifications	15

IDN

Technical specifications

1. Foreword

1.1. Reference documents

The implementation of IDNs at AFNIC is based on the **IDNA2008** standard, and the following reference documents.

- Definitions and protocol:
 - [RFC 5890 \(08/2010 23 pages\)](#): *Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework*
 - [RFC 5891 \(08/2010 17 pages\)](#): *Internationalized Domain Names in Applications (IDNA): Protocol*
 - [RFC 5892 \(08/2010 70 pages\)](#): *The Unicode Code Points and Internationalized Domain Names for Applications (IDNA)*
 - [RFC 5894 \(08/2010 43 pages\)](#): *Internationalized Domain Names for Applications (IDNA): Background, Explanation, and Rationale*
- Punycode encoding algorithm:
 - [RFC 3492 \(03/2003 35 pages\)](#): *Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)*

1.2. Decision of the Board

The open-door policy and the registration rules related to IDNs were approved by the AFNIC Board of Directors on 10 June 2011 and are available here:

<http://www.afnic.fr/medias/documents/afnic-politique-ouverture-idn.pdf>

1.3. Terms subject to prior review

The management of the terms subjected to prior review is not yet completely defined and will be the object of a later communication.

1.4. Brief backgrounder on IDN technology

The DNS protocol was not originally defined to be restricted to a set of characters. It is its use and other limitations of "the age" (the protocol is 30 years old) that have resulted in the definition of the syntactic rules we know today.

The purpose of the **IDNA2008** standard is to reconcile human needs and technical constraints by allowing the use of all forms of writing in domain names. All these forms of writing and the characters they use are defined and grouped together under a standard called Unicode. Since the syntactic rules for domain names require the use of single letters of the Latin alphabet ("a" to "z"), as well as numbers, hyphens, and periods to separate labels, a mechanism for the canonical formation of Unicode domain names and for encoding them has been developed to create names consistent with these rules. While in applications such as web browsers, Unicode names will be displayed, their DNS resolution will be performed using their encoded form (this is normally transparent to the user who should not have to handle this type of domain name).

1.5. *Warning*

Although its impact may seem small, it is important to note that AFNIC implements the **IDNA2008** standard, which slightly differs from the **IDNA2003** standard. With respect to the processing of the characters included, the German Eszett (ß) is encoded, not transformed into "ss" as in the previous version of the IDN standard. In addition, the canonicalization step (nameprep) has disappeared, which will have some impact on the use of our interfaces.

Each AFNIC application is now free to apply its own rules in this respect. Besides the fact that Unicode domain names must be in Normal Form C, we have chosen to allow the entry of capitals (to ensure backward compatibility with current uses) but their lower-case equivalents will actually be taken into account by the system (note that the Eszett is only accepted in its lower-case form). For example, the domain name "Thé-ou-Café.fr" is not legal in accordance with the **IDNA2008** standard. We shall accept it, however once it has been standardized as "thé-ou-café.fr".

With more "exotic" alphabets than the Latin, the problem will no doubt be more complex, but as long as AFNIC continues to use the characters indicated in the list below in this document, their canonical form will continue to apply.

1.6. Terms and definitions

- **Unicode:** Standard enabling any character in any form of writing to be encoded in a unique fashion ([Unicode on Wikipedia](#)).
- **UTF-8:** One of the encoding formats used to encode Unicode characters.
- **ISO-8859-15:** One of the ISO 8-bit encoding standards of the Latin alphabet. Also known as latin9.
- **LatinX:** Other names of certain ISO standards. Unlike Latin1, Latin9 includes the ligation "e in o".
- **LDH:** "LETTER-DIGIT-HYPHEN" the only ASCII characters authorized for the composition of a label in a domain name.
- **ASCII:** "American Standard Code for Information Interchange", the oldest computer standard for encoding characters. Strictly speaking 7-bit, it can only encode 128 characters.
- **ACE:** "ASCII Compatible Encoding" is the encoded version of a domain name in its LDH form (xn-caf-dma in Punycode, i.e. its "A-label form").
- **IDN:** "Internationalized Domain Name", containing characters other than ASCII characters alone.
- **Canonicalization:** The canonical formation of a string of characters. For example, in Latin, putting a string of characters in their lower-case form is one of the operations that can be involved in a canonicalization process.
- **Normal Form C:** Normal form requiring that the characters be (pre)composed. A character corresponds to a unique code point. This exclude characters obtained by using diacritical marks combined with base characters.
- **Code point:** Single index associated with a character.
- **Glyph:** Graphical representation of a character
- **NAMEPREP:** Defines the version in canonical form of a Unicode domain name (was part of IDNA2003, no longer exists in IDNA2008).
- **Punycode:** Reversible and unique algorithm, used to transform a canonicalized IDN into its ACE form.
- **Sunrise:** Period during which the registration of certain domain names is subject to certain conditions.
- **Grandfathering:** the rule used during the "sunrise" period to determine who may or may not file IDNs.

2. Table of accepted characters

The following table represents the set of characters may be used to compose the label of a domain name. Historically, only the first 37 characters in this table were allowed, **but as of May 3, 2012, it will be possible to use 30 new characters in the composition of the labels of domain names.** The "ASCII equivalent" column is special in that it will only be meaningful during the sunrise period (note that sometimes the ASCII equivalent of a Unicode character is a group of two characters). This will be detailed a little later in this document.

#	Code point	Glyph	Name	ASCII equivalent
1	U+002D	-	HYPHEN-MINUS SIGN	-
2	U+0030	0	DIGIT ZERO	0
3	U+0031	1	DIGIT ONE	1
4	U+0032	2	DIGIT TWO	2
5	U+0033	3	DIGIT THREE	3
6	U+0034	4	DIGIT FOUR	4
7	U+0035	5	DIGIT FIVE	5
8	U+0036	6	DIGIT SIX	6
9	U+0037	7	DIGIT SEVEN	7
10	U+0038	8	DIGIT EIGHT	8
11	U+0039	9	DIGIT NINE	9
12	U+0061	a	LATIN SMALL LETTER A	a
13	U+0062	b	LATIN SMALL LETTER B	b
14	U+0063	c	LATIN SMALL LETTER C	c
15	U+0064	d	LATIN SMALL LETTER D	d
16	U+0065	e	LATIN SMALL LETTER E	e
17	U+0066	f	LATIN SMALL LETTER F	f
18	U+0067	g	LATIN SMALL LETTER G	g
19	U+0068	h	LATIN SMALL LETTER H	h
20	U+0069	i	LATIN SMALL LETTER I	i
21	U+006A	j	LATIN SMALL LETTER J	j
22	U+006B	k	LATIN SMALL LETTER K	k
23	U+006C	l	LATIN SMALL LETTER L	l
24	U+006D	m	LATIN SMALL LETTER M	m
25	U+006E	n	LATIN SMALL LETTER N	n
26	U+006F	o	LATIN SMALL LETTER O	o
27	U+0070	p	LATIN SMALL LETTER P	p
28	U+0071	q	LATIN SMALL LETTER Q	q
29	U+0072	r	LATIN SMALL LETTER R	r
30	U+0073	s	LATIN SMALL LETTER S	s

31	U+0074	t	LATIN SMALL LETTER T	t
32	U+0075	u	LATIN SMALL LETTER U	u
33	U+0076	v	LATIN SMALL LETTER V	v
34	U+0077	w	LATIN SMALL LETTER W	w
35	U+0078	x	LATIN SMALL LETTER X	x
36	U+0079	y	LATIN SMALL LETTER Y	y
37	U+007A	z	LATIN SMALL LETTER Z	z
38	U+00DF	ß	LATIN SMALL LETTER SHARP S	ss
39	U+00E0	à	LATIN SMALL LETTER A WITH GRAVE	a
40	U+00E1	á	LATIN SMALL LETTER A WITH ACUTE	a
41	U+00E2	â	LATIN SMALL LETTER A WITH CIRCUMFLEX	a
42	U+00E3	ã	LATIN SMALL LETTER A WITH TILDE	a
43	U+00E4	ä	LATIN SMALL LETTER A WITH DIAERESIS	a
44	U+00E5	å	LATIN SMALL LETTER A WITH RING ABOVE	a
45	U+00E6	æ	LATIN SMALL LETTER AE	ae
46	U+00E7	ç	LATIN SMALL LETTER C WITH CEDILLA	c
47	U+00E8	è	LATIN SMALL LETTER E WITH GRAVE	e
48	U+00E9	é	LATIN SMALL LETTER E WITH ACUTE	e
49	U+00EA	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX	e
50	U+00EB	ë	LATIN SMALL LETTER E WITH DIAERESIS	e
51	U+00EC	ì	LATIN SMALL LETTER I WITH GRAVE	i
52	U+00ED	í	LATIN SMALL LETTER I WITH ACUTE	i
53	U+00EE	î	LATIN SMALL LETTER I WITH CIRCUMFLEX	i
54	U+00EF	ï	LATIN SMALL LETTER I WITH DIAERESIS	i
55	U+00F1	ñ	LATIN SMALL LETTER N WITH TILDE	n
56	U+00F2	ò	LATIN SMALL LETTER O WITH GRAVE	o
57	U+00F3	ó	LATIN SMALL LETTER O WITH ACUTE	o
58	U+00F4	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX	o
59	U+00F5	õ	LATIN SMALL LETTER O WITH TILDE	o
60	U+00F6	ö	LATIN SMALL LETTER O WITH DIAERESIS	o
61	U+00F9	ù	LATIN SMALL LETTER U WITH GRAVE	u
62	U+00FA	ú	LATIN SMALL LETTER U WITH ACUTE	u
63	U+00FB	û	LATIN SMALL LETTER U WITH CIRCUMFLEX	u
64	U+00FC	ü	LATIN SMALL LETTER U WITH DIAERESIS	u
65	U+00FD	ý	LATIN SMALL LETTER Y WITH ACUTE	y
66	U+00FF	ÿ	LATIN SMALL LETTER Y WITH DIAERESIS	y
67	U+0153	œ	LATIN SMALL LIGATURE OE	oe

3. Use of Unicode versions vs. LDH versions

Domain names are present in the server names, in the URL, and in the email addresses: here are the forms accepted by AFNIC interfaces. Detailed error messages will be returned in cases of non-compliance with these rules.

- **Domain name:** According to the interfaces, will be accepted in their Unicode version, in their LDH version, or 2.
- **Server name:** **ONLY** the LDH version is acceptable.
- **URL:** **ONLY** the LDH version is acceptable.
- **E-Mail:** **ONLY** the LDH version is acceptable.

4. Grandfathering Rule

As of July 3, 2012 (the exact schedule will be announced later), it will be possible for anyone to file an IDN as long as it is consistent with our naming policy where the rule "first come, first served" applies. On the other hand, over the preceding two months (**from May 3 to July 3**), a "sunrise" period with the "Grandfathering" rule will be established so that holders of domain names corresponding to the current rules have priority in reserving an IDN corresponding to these ASCII versions. This is where the concept of ASCII equivalence mentioned above comes into play.

What exactly does this mean? Quite simply, the entity filing an IDN must hold an ASCII-equivalent domain name for the transaction to be accepted. However, if the ASCII equivalent has not yet been filed, it will be possible for the holder to create it at any time, even during the sunrise period. Here are a few examples of potential filings during the sunrise.

Caution: the concept of an ASCII-equivalent domain name is applied to the entire domain name (including the extension). A holder of "café.fr" is entitled to register "café.fr", but not "café.re".

4.1. Example 1

IDN	m ü l l e r - s t r a ß e . f r
ASCII equivalent	m u l l e r - s t r a s s e . f r

In this case only the holder of the domain name "muller-strasse.fr" may file "müller-straße.fr" (whose ACE form is xn--mller-strae-46a18a.fr). The holder can also equally well register "muller-straße.fr" (whose ACE form is xn--muller-strae-v9a.fr) or even "müller-strasse.fr" (whose ACE form is xn--mller-strasse-wob.fr) since these three IDNs share the same ASCII-equivalent form.

4.2. Example 2

IDN 1	p	ê	c	h	é	s	.	f	r
IDN 2	p	é	c	h	é	s	.	f	r
ASCII equivalent	p	e	c	h	e	s	.	f	r

In this second example, we have two IDNs that share the same ASCII-equivalent version. As in the previous example, only the current holder may file one or both forms (or all the forms sharing the same ASCII-equivalent version) during the sunrise period. On the other hand, if, in our example, the owner of "peche.fr" only registers "pêchés.fr" (whose ACE form is xn--pchs-dpad.fr) during the sunrise, at the end of the sunrise, the holder will have no claim to "péchés.fr" (whose ACE form is xn--pchs-bpac.fr) which can then be registered by a different holder.

5. EPP specifications

With regard to EPP, the only acceptable form for domain names is the LDH form, i.e. the ACE version for IDNs.

Some registries require the use of an extension in order to indicate the "language / script tag" to be considered for processing the order. Some rules (lists of characters included, rules for canonical form, etc.) may differ from one language / form of writing to another. Normally the extension is one required to process the IDNs. However, two reasons have led us not to implement this change (for the time being).

- In the current situation, it does not make much sense for AFNIC, since the same rule and same list of characters will apply for all the ccTLDs we currently manage. In addition, we only accept characters in the Latin alphabet. Things might have been different if we had accepted, in addition to this alphabet, the Cyrillic or Arabic alphabet, but that is not the case.
- There is no official EPP extension to manage this configuration, and with work related to the introduction of the new gTLDs, the need has become apparent, and the discussions within the former IETF working groups that led to the development of the EPP protocol have restarted. So it seems better to wait rather than to create an EPP extension specific to AFNIC.

To avoid the risk of implementing an EPP extension that may not match the consensus that might emerge from current discussions, when there is no real need for AFNIC to do so, we have decided to wait and see. If a standard emerges, however, we have not excluded the possibility of implementing an EPP extension. But if this were to happen, in any case it would be after the launch of IDNs, and after consulting you in order to determine which would be the best schedule for the upgrade.

So for EPP, apart from the fact that the domain names and servers are only accepted in their LDH form, no change is to be expected.

5.1. Example of an IDN query using EPP

```
C:<epp xmlns="urn:ietf:params:xml:ns:epp-1.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:ietf:params:xml:ns:epp-1.0 epp-1.0.xsd">
C: <command>
C:   <info>
C:     <domain:info xmlns:domain="urn:ietf:params:xml:ns:domain-1.0"
xsi:schemaLocation="urn:ietf:params:xml:ns:domain-1.0 domain-1.0.xsd">
C:       <domain:name hosts="all">xn--strae-42-tya.fr</domain:name>
C:     </domain:info>
C:   </info>
C:   <clTRID>PasTerribleCommeSecret666</clTRID>
C: </command>
C:</epp>
```

5.2. Example of response

```
S:<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
S:<epp xmlns="urn:ietf:params:xml:ns:epp-1.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:ietf:params:xml:ns:epp-1.0 epp-1.0.xsd">
S: <response>
S:   <result code="1000">
S:     <msg>Command completed successfully</msg>
S:   </result>
S:   <resData>
S:     <domain:infData xmlns:domain="urn:ietf:params:xml:ns:domain-1.0">
S:       <domain:name>xn--strae-42-tya.fr</domain:name>
S:       <domain:roid>DOM000003382455-FRNIC</domain:roid>
S:       <domain:status s="inactive"/>
S:       <domain:registrant>TGCA108</domain:registrant>
S:       <domain:contact type="admin">TGCA108</domain:contact>
S:       <domain:contact type="tech">VL0</domain:contact>
S:       <domain:clID>-naqjanir485-.fr</domain:clID>
S:       <domain:crDate>2012-01-20T13:16:24.0Z</domain:crDate>
S:       <domain:exDate>2013-01-20T00:00:00.0Z</domain:exDate>
S:       <domain:upDate>2012-01-20T13:16:24.0Z</domain:upDate>
S:       <domain:authInfo>
S:         <domain:pw>IDN2012</domain:pw>
S:       </domain:authInfo>
S:     </domain:infData>
S:   </resData>
S:   <trID>
S:     <clTRID>PasTerribleCommeSecret666</clTRID>
S:     <svTRID>DEV-vraiton-16996-14-1327418457.36048</svTRID>
S:   </trID>
S: </response>
S:</epp>
```

6. Web Interface Specifications

The forms used for the administration of domain names will be changed so that an encoding step with validation of the two forms for IDNs is systematically performed. It will be possible to access these forms by entering one or other of the forms (Unicode vs. ASCII), but a validation step with visualization of the domain name in its two forms has been added to avoid any ambiguity about the domain name you want to process. In addition, in order to avoid any encoding problems during copy/paste operations between applications/systems, a virtual keyboard can be used to capture additional characters, some of which (e.g. the German eszett) are not available on the keyboards of every user.

We also plan to make available a tool for converting IDNs on all the pages of our site.

7. DAS specifications

Accounting for IDNs is an integral part of the existing DAS protocol used by AFNIC, namely **IRIS:DCHK** ([RFC 5144](#)); on the other hand, this protocol refers to the **IDNA2003** standard. In the DAS protocol implemented by AFNIC (**IDNA2008**), the Nameprep step no longer exists. However, as discussed above, given the alphabet used, and in order to be consistent with our other interfaces and current uses at AFNIC, we also accept uppercase input.

It will be possible to query an IDN in its ASCII or Unicode form; on the other hand, the 'entityClass' attribute of the <lookupEntity> element will not be the same. In the case of an ASCII form, indicate "domain-name"; in the case of a Unicode form, indicate "idn". This is not specific to AFNIC, so if your client code complies with the RFC, no change is to be expected.

The answer, in the case of an IDN query, will contain an additional element, namely <idn> containing the Unicode version of the domain name. On the other hand, unlike the domain name entered, only the 67 characters listed above will be used as output (no capitals). The <domainName> element of the answer will always contain the ASCII form of the domain name. The values of the 'entityClass' and 'entityName' attributes in the answer will be identical to those in the query.

7.1. Example of an IDN query in its Unicode form

Query:

```
<?xml version="1.0" encoding="UTF-8"?>
<iris1:request xmlns:iris1="urn:ietf:params:xml:ns:iris1">
  <iris1:searchSet>
    <iris1:lookupEntity registryType="dchk1" entityClass="idn"
entityName="STraße-42.fr"/>
  </iris1:searchSet>
</iris1:request>
```

Answer:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<iris:response xmlns:iris="urn:ietf:params:xml:ns:iris1">
  <iris:resultSet>
    <iris:answer>
      <domain xmlns="urn:ietf:params:xml:ns:dchkl" authority="fr"
registryType="dchkl" entityClass="idn" entityName="STraße-42.fr">
        <domainName>xn--strae-42-tya.fr</domainName>
        <idn>straße-42.fr</idn>
        <status>
          <inactive/>
        </status>
        <createdDateTime>2012-01-20T13:16:24.0Z</createdDateTime>
        <lastDatabaseUpdateDateTime>2012-01-
20T13:16:24.0Z</lastDatabaseUpdateDateTime>
      </domain>
    </iris:answer>
  </iris:resultSet>
</iris:response>
```

7.2. Example of an IDN query in its ACE form**Query:**

```
<?xml version="1.0" encoding="UTF-8"?>
<iris1:request xmlns:iris1="urn:ietf:params:xml:ns:iris1">
  <iris1:searchSet>
    <iris1:lookupEntity registryType="dchkl" entityClass="domain-
name" entityName="xn--strae-42-tya.fr"/>
  </iris1:searchSet>
</iris1:request>
```

Answer:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<iris:response xmlns:iris="urn:ietf:params:xml:ns:iris1">
  <iris:resultSet>
    <iris:answer>
      <domain xmlns="urn:ietf:params:xml:ns:dchkl" authority="fr"
registryType="dchkl" entityClass="domain-name" entityName="xn--
strae-42-tya.fr">
        <domainName>xn--strae-42-tya.fr</domainName>
        <idn>straße-42.fr</idn>
        <status>
          <inactive/>
        </status>
        <createdDateTime>2012-01-20T13:16:24.0Z</createdDateTime>
        <lastDatabaseUpdateDateTime>2012-01-
20T13:16:24.0Z</lastDatabaseUpdateDateTime>
      </domain>
    </iris:answer>
  </iris:resultSet>
</iris:response>
```

7.3. Example of an IDN query in its Unicode form during the sunrise period

Query:

```
<?xml version="1.0" encoding="UTF-8"?>
<iris1:request xmlns:iris1="urn:ietf:params:xml:ns:iris1">
  <iris1:searchSet>
    <iris1:lookupEntity registryType="dchk1" entityClass="idn"
entityName="àfnic.fr"/>
  </iris1:searchSet>
</iris1:request>
```

Answer:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<iris:response xmlns:iris="urn:ietf:params:xml:ns:iris1">
  <iris:resultSet>
    <iris:answer>
      <domain xmlns="urn:ietf:params:xml:ns:dchk1" authority="fr"
registryType="dchk1" entityClass="idn" entityName="àfnic.fr">
        <domainName>xn--fnic-zna.fr</domainName>
        <idn>àfnic.fr</idn>
        <status>
          <reserved>
            <subStatus
authority="fr">grandfathering_rule</subStatus>
              <description language="en">ASCII equivalent domain
already exists</description>
            </reserved>
          </status>
        </domain>
      </iris:answer>
    </iris:resultSet>
  </iris:response>
```

8. Whois port 43 Specifications

It will be possible to query the Whois server in 2 ways for IDNs: either by using the Unicode form or by using the ACE form. We also plan to offer an option to control the output format (UTF-8 vs. ASCII). We have chosen to develop this interface by default, and shall provide UTF-8.

Two new fields will appear in the Whois output. The first 'domain-ace:' will be systematically present in the case of IDNs, regardless of the chosen output format. The second 'domain-idn:' will only be present in the case of an UTF-8 output (the default output). The 'domain:' field uses the form used in the query. Note that for non-IDN domain names, there is no change. Also take care about being consistent in suing options: for example, it will not be possible to query a domain name in its Unicode form if you ask for an ASCII output format.

Here is a summary table of the presence of these new fields depending on the context when you query an IDN.

Output format	IDN input format	'domain:' field format	Presence 'domain-ace:' field	Presence 'domain-idn:' field
UTF-8	Unicode	Unicode	X	X
UTF-8	ASCII	ASCII	X	X
ASCII	Unicode	Error	-	-
ASCII	ASCII	ASCII	X	-

Example of an IDN query (the request was for 'straße.fr') in the UTF-8 output format (default value).

```

domain:      straÙe.fr
domain-ace:  xn--strae-oqa.fr
domain-idn:  straÙe.fr
status:     ACTIVE
hold:       NO
holder-c:   AFNI21-FRNIC
admin-c:    NFC1-FRNIC
tech-c:     NFC1-FRNIC
zone-c:     NFC1-FRNIC
nsl-id:     NSL53707-FRNIC
dsl-id:     SIGN6-FRNIC
registrar:  AFNIC registry
anniversary: 01/01
created:    01/01/1995
last-update: 28/09/2011
source:     FRNIC

```

The exact value of the output format control options will be announced later.

9. Notification e-mail specifications

The notifications e-mail specifications will be communicated later.