The Italian approach to Internationalized Domain Names (IDNs)

Dr. Domenico Laforenza
Director of the Institute for Informatics and Telematics
Italian National Research Council, Pisa, Italy

W3C Workshop
Content on the Multilingual Web
4 - 5 April 2011, Pisa, Italy
OUTLINE

- The Internet Ecosystem
- Domain Name System
- gTLDs and ccTLDs
- .it Registry
- Internationalized Domain Names (IDNs)
- IDNs under .it
The Domain Name System

This photo was published in the August 8, 1994 issue of Newsweek and commemorates the 25th anniversary of the ARPANET. Jon Postel, Steve Crocker and Such was the state of networking in the primitive 1960s...

Picture from Vint Cerf
The Domain Name System

• Every TCP/IP resource can be assigned with a name (*symbolic name*)
• A mechanism to associate the symbolic name of a host to its IP address(es): *direct resolution*
• A mechanism to associate the IP address of a host to its symbolic name: *inverse resolution*

• **Domain Name System (DNS)**
  – Defined by ISI - USC 1984
  – RFC 882, RFC 883, RFC 973 (obsolete)
  – RFC 1034, RFC 1035, RFC 1123, RFC 1537, RFC 1912, RFC 2182 and their updates
The namespace

- Namespace organized according to a hierarchical model:
  - the DNS database logical structure is like a “overturned tree”
  - Each tree node represents a domain
  - Each domain can have children: the sub-domains
  - Each node is identified by a label

The root of the tree is unique and its label is empty. Generally the root is identified by a “.".

- Namespace structure (RFC 1591):
  - general Top Level Domains (gTLDs)
  - country code Top Level Domains (ccTLDs)
  - Domains for the inverse resolution (arpa)
DNS Logical Structure

1st Level Domains:
- edu
- com
- org
- it

2nd Level Domains:
- stanford
- mit
- ibm
- ferrari
- w3c
- cnr
The first level domains are classified in gTLDs and ccTLDs

Historical generic Top Level Domains:

- **com**: commercial organizations
- **edu**: USA university and research
- **gov**: USA government organizations
- **mil**: USA military organizations
- **net**: open to everyone
- **org**: open to everyone
- **int**: international organizations, treaties

http://www.w3c.org
gTLDs

- **aero**
  - for airline companies - Societe Internationale de Telecommunications Aeronautiques SC, (SITA (http://www.nic.aero)

- **biz**
  - for companies - NeuLevel, Inc. (http://www.nic.biz)

- **info**
  - “unrestricted use” - Afilias Lt (http://www.nic.info)

- **museum**
  - for museums - Museum Domain Management Association, (MDMA) (http://musedoma.museum)

- **name**
  - for individuals - Global Name Registry, Ltd (http://www.name)

- **pro**
  - for freelances - RegistryPro, Ltd (http://www.registry.pro)
gTLDs

- **jobs**
  - for companies dealing with human resources management - Employ Media LLC: http://www.goto.jobs

- **mobi**
  - for mobile and related services companies - mTLD Top Level Domain, Ltd.: http://pc.mtld.mobi

- **travel**
  - for travel agencies - Tralliance Corporation (http://www.tralliance.info)

- **cat**
  - for the Catalan cultural and linguistic community - Fundació puntCat: http://www.domini.cat

- **xxx**
  - for pornographic sites (ICANN approved on 18 March 2011)
A country code Top-Level Domain is an Internet top-level domain used or reserved for a country, a sovereign state, or a dependent territory.

National domains, represented by the ISO 3166-1 codes:

- .al Albania
- .de Germany
- .it Italy
- .uk Great Britain

--------
DNS Query: An example

Domain Name Resolution Process
ABROAD .IT ON THE UP

Demand for dot it domains is on the up and up, even abroad.

Growth in the demand for dot it domains: in the space of a years the numer of dot it domains registered by companies outside Italy has more than doubled.

Download Focus it
Pisa, October 22nd, 2010

**TOP CCTLD REGISTRIES BY DOMAIN NAME BASE, SECOND QUARTER 2010**

1. .de (Germany)       6. .ru (Russian Federation)
2. .uk (United Kingdom) 7. .br (Brazil)
3. .cn (China)         8. .ar (Argentina)
4. .nl (Netherlands)    9. .it (Italy)
5. .eu (European Union) 10. .tk (Tokelau)

Source: Zooknic, August 2010
IIT-CNR: The Mission

• Carry out research, development, technology transfer and training in the fields of information and communication technologies and computational science

• In line with the strategy of the:
  ❖ CNR ICT Department
  ❖ EU FP7

• IIT is naturally and concretely projected towards the Internet of the Future
The Institute for Informatics and Telematics (IIT-CNR)

Research Groups

- Algorithms and Computational Mathematics
- Trust, Security and Privacy in Internet
- Ubiquitous Internet
- Web Applications for the Future Internet
- Marco Conti
- Andrea Marchetti
- Paola Favati
- Paolo Mori
The Institute for Informatics and Telematics (IIT-CNR)
Technology Services

- Planning, Design, and Monitoring of Telematic Networks
- Technology Services
- Internet Services and Technological Development

Maurizio Martinelli
Marco Sommani
Domenico Laforenza
The .it Registry: a brief overview

- Delegated to CNR on December 23rd, 1987
- More than 2,140,000 domain names
- New “synchronous” (Extensible Provisioning Protocol - EPP) registration system from September 28th, 2009
- Coexistence of the “asynchronous” and “synchronous” systems until August 2011
- About 1,200 Registrars
- Open to EU juridical and natural persons
The .it Registry: organizational structure

Management Committee

Director IIT/Registry

Rules Committee

Operations

Legal

Systems & Development

External Relations & Communication

International Relations (pro tempore)

About 70 people including staff-operators, administrative persons and technicians
<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>Internet Users</th>
<th>P. R.</th>
<th>% Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>3,834,792,852</td>
<td>872,526,978</td>
<td>22.8%</td>
<td>43.0%</td>
</tr>
<tr>
<td>Europe</td>
<td>813,319,511</td>
<td>475,123,735</td>
<td>58.4%</td>
<td>23.4%</td>
</tr>
<tr>
<td>North America</td>
<td>344,124,450</td>
<td>271,330,900</td>
<td>78.8%</td>
<td>13.4%</td>
</tr>
<tr>
<td>Latam / Carib.</td>
<td>592,556,972</td>
<td>209,874,973</td>
<td>35.4%</td>
<td>10.3%</td>
</tr>
<tr>
<td>Africa</td>
<td>1,013,779,050</td>
<td>115,631,340</td>
<td>11.4%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Middle East</td>
<td>212,336,924</td>
<td>63,708,386</td>
<td>30.0%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Oceania, Aust.</td>
<td>34,700,201</td>
<td>21,272,470</td>
<td>61.3%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Total World</td>
<td>6,845,609,960</td>
<td>2,029,468,782</td>
<td>29.6%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Source: Internet World Stats, preliminary estimates for December 31, 2010. Notes: a) Population data comes from the US Census Bureau; b) Internet user estimates come from ITU, Nielsen Online, GfK, official regulating agencies and other trustworthy surveys; c) Mexico is included in Central America and Turkey is included in Europe, according to the United Nations country classification; d) P.R. means the Penetration Rate (Internet users times 100 divided by population). Copyright © 2011, Miniwatts Marketing Group.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>8,214,160</td>
<td>6,143,600</td>
<td>74.8%</td>
<td>192.6%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Belgium</td>
<td>10,423,473</td>
<td>8,113,200</td>
<td>77.8%</td>
<td>305.7%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>7,148,785</td>
<td>3,395,000</td>
<td>47.5%</td>
<td>689.5%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Cyprus</td>
<td>1,102,677</td>
<td>433,800</td>
<td>39.3%</td>
<td>261.5%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>10,201,707</td>
<td>6,680,800</td>
<td>65.5%</td>
<td>568.1%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Denmark</td>
<td>5,515,575</td>
<td>4,750,500</td>
<td>86.1%</td>
<td>143.6%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Estonia</td>
<td>1,291,170</td>
<td>969,700</td>
<td>75.1%</td>
<td>164.5%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Finland</td>
<td>5,255,695</td>
<td>4,480,900</td>
<td>85.3%</td>
<td>132.5%</td>
<td>0.9%</td>
</tr>
<tr>
<td>France</td>
<td>64,768,369</td>
<td>44,625,300</td>
<td>68.9%</td>
<td>425.0%</td>
<td>9.4%</td>
</tr>
<tr>
<td>Germany</td>
<td>82,282,988</td>
<td>65,123,800</td>
<td>79.1%</td>
<td>171.3%</td>
<td>13.7%</td>
</tr>
<tr>
<td>Greece</td>
<td>10,749,943</td>
<td>4,970,700</td>
<td>46.2%</td>
<td>397.1%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Hungary</td>
<td>9,992,339</td>
<td>6,176,400</td>
<td>61.8%</td>
<td>763.8%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Ireland</td>
<td>4,622,917</td>
<td>3,042,600</td>
<td>65.8%</td>
<td>288.1%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Italy</td>
<td>58,090,681</td>
<td>30,026,400</td>
<td>51.7%</td>
<td>127.5%</td>
<td>8.3%</td>
</tr>
<tr>
<td>Latvia</td>
<td>2,217,869</td>
<td>1,503,400</td>
<td>67.6%</td>
<td>902.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Lithuania</td>
<td>3,545,319</td>
<td>2,103,471</td>
<td>59.3%</td>
<td>834.9%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>497,538</td>
<td>424,500</td>
<td>85.3%</td>
<td>324.5%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Malta</td>
<td>406,771</td>
<td>240,600</td>
<td>59.1%</td>
<td>501.5%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>16,783,092</td>
<td>14,872,200</td>
<td>88.6%</td>
<td>281.3%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Poland</td>
<td>38,463,689</td>
<td>22,450,600</td>
<td>58.4%</td>
<td>701.8%</td>
<td>4.7%</td>
</tr>
<tr>
<td>Portugal</td>
<td>10,736,765</td>
<td>5,168,800</td>
<td>48.1%</td>
<td>106.8%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Romania</td>
<td>21,959,278</td>
<td>7,786,700</td>
<td>35.5%</td>
<td>873.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Slovakia</td>
<td>5,470,306</td>
<td>4,063,600</td>
<td>74.3%</td>
<td>525.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Slovenia</td>
<td>2,005,692</td>
<td>2,003,136</td>
<td>64.8%</td>
<td>332.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Spain</td>
<td>46,505,963</td>
<td>29,093,984</td>
<td>62.6%</td>
<td>440.0%</td>
<td>8.1%</td>
</tr>
<tr>
<td>Sweden</td>
<td>9,074,055</td>
<td>3,397,900</td>
<td>38.2%</td>
<td>107.5%</td>
<td>1.8%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>62,348,447</td>
<td>51,442,100</td>
<td>82.5%</td>
<td>234.0%</td>
<td>10.8%</td>
</tr>
<tr>
<td><strong>European Union</strong></td>
<td><strong>499,671,847</strong></td>
<td><strong>337,779,055</strong></td>
<td><strong>67.6%</strong></td>
<td><strong>257.8%</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

**NOTES:** (1) The European Union Internet Statistics were updated for June 30, 2010. (2) Population is based on data from the Census Bureau. (3) The usage numbers come from various sources, mainly from data published by Nielsen Online, ITU, GfK, local NICs and private sources. (4) Data may be cited, giving due credit and establishing an active link to Internet World Stats. Copyright © 2010, Miniwatts Marketing Group. All rights reserved.
Today’s Internet user is increasingly likely to be a non-English speaker who does not use Latin script characters in every day life.

- **60% of Internet users are non-English speakers**, while the dominant language used on the Internet is English. Therefore, it is important to depict domain names in **non-Latin scripts**, called **Internationalized Domain Names (IDNs)**, to enhance the ability of all users to use the Internet in their own languages.

http://www.icann.org/en/topics/idn/
Script for everyone.....
(http://www.unicode.org/charts/#scripts)
Script for everyone.....
(http://www.unicode.org/charts/#scripts)
Internationalized Domain Names (IDNs)

- IDNs are domain names represented by local language characters (non-ASCII characters). Such domain names could contain characters with diacritical marks (ñ, é) as required by many European languages, or characters from non-Latin scripts (for example, Arabic or Chinese)
  - are domain names with characters other than a, b..., z; 0, 1..., 9; and “-“
- IDNs on the second and third levels exist in some gTLDs and in some ccTLDs

- TLD Registrars determine the choice of characters available under these TLDs (see the Italian approach later on)
IDN TLDs will be made available through two separate processes, initially through the New gTLD Program and the IDN ccTLD Fast Track Process.

IDN SLDs Usually a reference for domain names with local characters at the Second Level, while the top level remains in ASCII-only characters.

- For example: [παράδειγμα.test] ("example.test" in Greek)

IDN TLDs Usually the short reference for internationalized Top Level labels referring to the entire domain name being represented by local characters.

- For example: [실례.테스트] ("example.test" in Hangul (Korean script))
The First Four

- Special recognition of the first four IDN ccTLD:
  - United Arab Emirates, Saudi-Arabia, Russian Federation, Egypt
  - simultaneous delegation: مصر and امارات السعودية
  - .рф followed shortly after

- Milestones include:
  - Jun 2003: IDNA protocol, IDN Guidelines, initiation browser implemented
  - Feb 2008: First report from IDNC leading to IDN ccTLD Fast Track Process
  - Oct 2008: First draft implementation plan for IDN ccTLD Process
  - Oct 2009: ICANN Board approves IDN ccTLD Fast Track Process
  - Nov 2009: Launch of IDN ccTLD Fast Track Process
  - Jan 2010: First four IDN ccTLD labels approved
  - Apr 2010: First four IDN ccTLD approved to be delegated
  - May 2010: Introduction of the first four
IDNs and DNS

- IDNs will impose minimal changes to current DNS system
- IDNs must not break existing structure and hierarchy
- What is the solution?

✔ Punycode (RFC3492)

- IDNs are stored in the Domain Name System as ASCII strings using Punycode transcription
Punycode (RFC3492)

- Sequence of ASCII characters all IDNs will be encoded into in order for the Domain Name System (DNS) to understand and manage the names.
  - The intention is that domain name registrants and users will never see this decoded form of a domain name.
- The sole purpose is for the DNS to be able to resolve for example a web address containing local characters.
- The DNS is only capable of handling ASCII characters.
- For example, the punycode version of "हिंदी", in Devanagari script, version of “example.test”) is:
  - xn--p1b-6ci4b4b3a.xn--11b5bs3a9aj6g
- The prefix for the Punycode version of the domain names is always “xn--”. Hence this prefix is often reserved at the registry level to avoid confusion in registration of IDNs.
IDNA – Protocol Functionality

- Domain Name Resolution Process:

  http://www.서울.test → xn--g2b8q.test
  → Local Server
  → IP address of xn--g2b8q.test
  → Root Server
  → .test Server
  → 서울.test Server

IDNA is a client-based protocol:
1. User types in 서울.test in for example a browser
2. 서울.test gets converted to Unicode,
   - if not already entered as such
3. IDNA conversion → xn--g2b8q.test

Domain Name Resolution Process
Internationalizing Domain Names in Applications (IDNA)

- A mechanism defined in 2003 for handling IDNs containing non-ASCII characters.
  - An IDNA-enabled application is able to convert between the internationalized and ASCII representations of a domain name.
  - It uses the ASCII form for DNS lookups but can present the internationalized form to users who presumably prefer to read and write domain names in non-ASCII scripts such as Arabic or Hiragana.
  - Applications that do not support IDNA will not be able to handle domain names with non-ASCII characters, but will still be able to access such domains if given the (usually rather cryptic) ASCII equivalent.
  - Although the Domain Name System supports non-ASCII characters, applications such as email and web browsers restrict the characters which can be used as domain names for purposes such as a hostname.
Why are IDNs important?

• The most important is the growing number of Internet users around the world for whom it is difficult to use ASCII characters.
  – The fact is, the Internet is accessed by more people who do not use Latin languages and scripts than those who do.
  • This means that it is difficult for them to recognize ASCII characters and reproduce them on keyboards or use software to enter website addresses in browsers.
Internationalised Domain Names

Registro .it

Valle D’Aosta
Trentino Alto-Adige
Friuli Venezia Giulia
Piemonte
Lombardia
Veneto
Emilia-Romagna
Liguria
Toscana
Umbria
Marche
Abruzzo
Lazio
Molise
Campania
Basilicata
Calabria
Sicilia
Sardegna

focus .it

we like .it!
il Registro di domani / via libera agli Ildn
e a gennaio arrivano
i domini con l’accento

di Stefania Fabbri

espressione complessa - ma con un connotato facile da ricordare - che indica il sistema tramite il quale si può “registrare” sul Internet in domane piuttosto che in cinese, utilizzando lettere accedute o caratteri non latini. Sul web, fino a poco tempo fa, la scelta dei nomi era limitata da caratteri dell'alfabeto latino impiegati in lingue inglesi (oltre ai dialetti letterali e ai tratto “.”). L’uso degli Ildn, introdotti da locccon (Internet corporation for assigned names and numbers) nel 2000 durante il meeting di Seoul, rappresentava uno svolto soprattutto per le popolazioni di lingua araba, cinese o indù; centinaia di milioni di utenti Internet del mondo che fino a ieri erano di fatto costretti a usare un alfabe-
to non proprio. Per questo riguardo l’Italia, l’impatto degli accenti sono sicuramente più contenuto ma dà a tutti le possibilità di registrare domini esattamente uguali al nome dello stesso, della società o del marchio commerciale pressato. Gli Ildn rappresentano sicuramente un vantaggio per le registra-
tori non italiani. Per quanto riguarda l’Italia, l’impatto degli accenti sono sicuramente più contenuto ma dà a tutti le possibilità di registrare domini esattamente uguali al nome dello stesso, della società o del marchio commerciale pressato.
<table>
<thead>
<tr>
<th>Regione</th>
<th>% domini</th>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trentino Alto Adige</td>
<td>2,54%</td>
<td>388,76</td>
</tr>
<tr>
<td>Lombardia</td>
<td>21,92%</td>
<td>351,04</td>
</tr>
<tr>
<td>Toscana</td>
<td>7,83%</td>
<td>330,70</td>
</tr>
<tr>
<td>Lazio</td>
<td>11,85%</td>
<td>328,01</td>
</tr>
<tr>
<td>Emilia Romagna</td>
<td>8,61%</td>
<td>308,26</td>
</tr>
<tr>
<td>Marche</td>
<td>2,93%</td>
<td>296,29</td>
</tr>
<tr>
<td>Veneto</td>
<td>9,03%</td>
<td>289,76</td>
</tr>
<tr>
<td>Umbria</td>
<td>1,58%</td>
<td>276,39</td>
</tr>
<tr>
<td>Valle d'Aosta</td>
<td>0,22%</td>
<td>272,84</td>
</tr>
<tr>
<td>Friuli V. G.</td>
<td>2,07%</td>
<td>265,17</td>
</tr>
</tbody>
</table>

Internet Diffusion in Italy: The Top 10 Regions (TP = Penetration Rate)

TP out of 10.000 inhabitants

French Language

German Language
• Most of the words and names, in the Italian language are represented by ASCII characters (a-z A-Z 0-9) and, optionally, hyphen (-)

• IDNs will be an advantage for bilingual Italian regions (Valle d’Aosta and Trentino Alto Adige), where they may need to register domains in French or German.
IDNs under .it

Examples

www.papa.it

• papa (The Pope)
  – From Greek: πάππας (pappas)
  – www.papa.it

• papà (The Father)
  – www.papà.it (www.xn--pap-cla.it)
IDNs: PROs

- Possibility of registering a domain name **exactly equal to the real name** of the person, company, brand, etc.:
  - giosuècarducci.it (xn--giosucarducc-1db)
  - cittàdicastello.it (xn--cittdicastell-sdb)
  - nestlè.it (xn--nestl-vqa)

- Allow the registration of domain names in languages other than English in order to cater to the needs of those Italian Regions where there is Italian bilingualism (Valle d’Aosta e Trentino Alto Adige)

- Reduction in the future, any disputes on domain names – **nestle.it** IS NOT EQUAL TO nestlè.it

- Increase in registrations of domain names
IDNs: CONs

• The introduction of IDNs could generate, in some cases, confusion to the end user:
  • www.agora.it e www.agorà.it
  • Web sites can target different

• Development and maintenance of many applications used in the Registry
IDNs under .it

1) Registration of the domain names in ACE (ASCII Compatible Encoding– es. xn--pap-cla.it) or IDN (es. papà.it) formats
   – ACE will be the format, however, necessary for the generation of the .it zone and of the geographic domain names zones

2) Simultaneous starting of IDNs for the Italian language, for the German and French
3) Definition of the Unicode character set to support:

- The Latin-1 supplement and Latin Extended-A sets are more than enough to support accented characters of Italian, French and German
  - characters to support in the Italian language:
    - à, è, é, ì, ò, ù
  - characters to support in the French language:
    - à, â, æ, ç, è, é, ê, î, ï, ò, û, ü, ÿ
  - characters to support in the German language:
    - ß, ä, ö, ü
• In essence, combining the non-ASCII characters of the three alphabets, we get the following set of **21 characters** (and similar capital letters):

• à, â, ä, è, é, ê, ē, ī, î, ì, ò, ô, ö, ù, û, ü, æ, œ, ç, ß, Ŷ

  – ß was introduced in the encoding Unicode (in July 2008) +1E9E LATIN CAPITAL LETTER SHARP S.
IDNs under .it

- Registration of the domain name in IDN format
- Whois query both in IDN and Punycode
- Output both in IDN and Punycode
- Authoritative nameservers for a domain name in the Punycode format
- Generation of DNS zones in Punycode format
IDNs under .it: Sunrise or Landrush?

• "Landrush" is a term used to define the stampede of registrations when a registry opens to registrations.

• We will adopt the Landrush procedure

• Landrush: Pros
  – First come first served
  – Total lack of organizational and management problems

• Landrush: Cons
  – Risk of cybersquatting
IDNs under .it

Test period

TEST

IDNs: Available on the test platform
January 2011

Push Me

IDNs Official Landrush
July 2011
ITALY is joining the rest of the IDNs Community

Creation of Adam (1511) by Michelangelo, Sistine Chapel, Vatican
Acknowledgments
• I would like to express my gratitude to Dr. Maurizio Martinelli, responsible for the Internet Services and Technological Development Department, IIT-CNR
Grazie

Dr. Domenico Laforenza
domenico.laforenza@iit.cnr.it