

.eu Insights

World report on Internationalised Domain Names 2014

August 2014



United Nations
Educational, Scientific and
Cultural Organization

With the support of
**Communication and
Information Sector**

With the support of



VERISIGN™

EURid

www.eurid.eu

With the support of CENTR, LACTLD, APTLD, and AFTLD

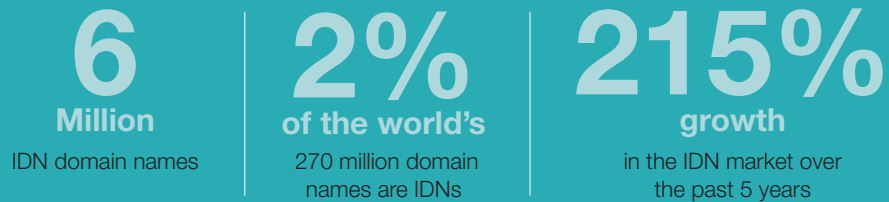
.eu Insights

The EURid Insights series aims to analyse specific aspects of the domain name environment. The reports are based on surveys, studies and research conducted by EURid in cooperation with industry experts and sector leaders.

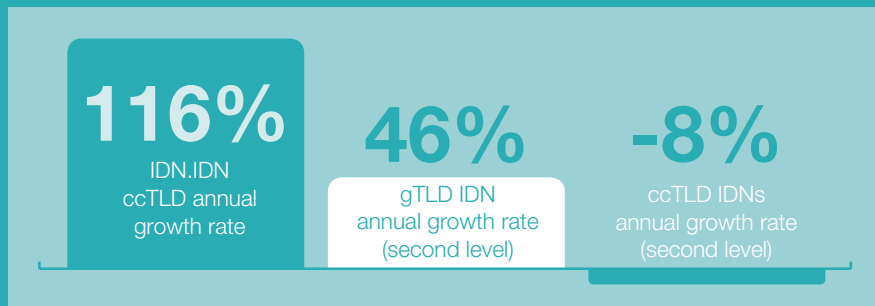


World report on Internationalised Domain Names 2014

Numbers (December 2013)



2012-2013



Languages

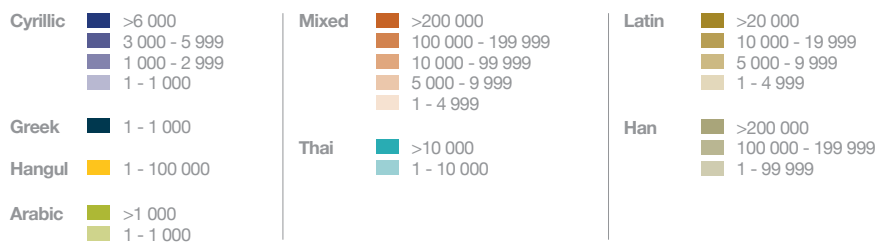
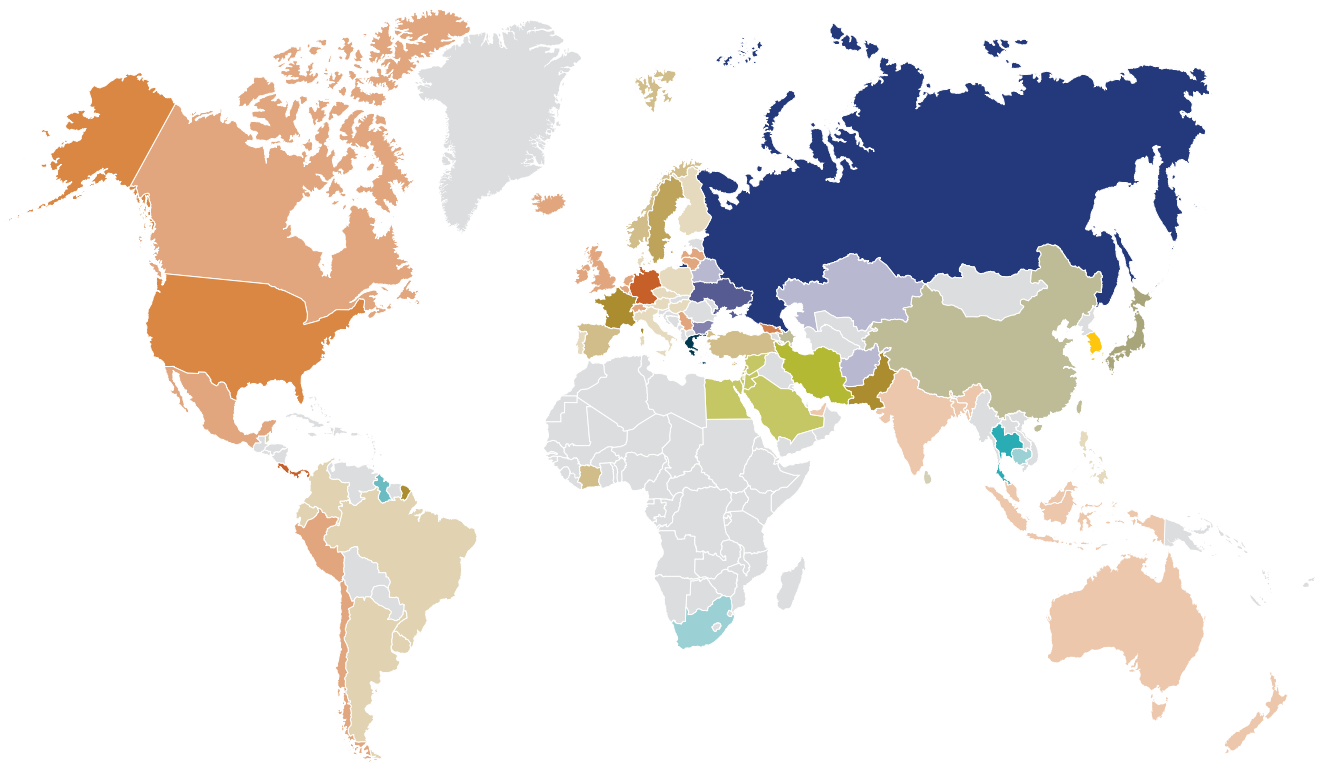
- IDNs help to enhance multilingualism in cyberspace
- The IDN market is more balanced in favour of emerging economies
- IDNs are accurate predictors of the language of online content
- 99%+ correlation between IDN scripts and language of website
- Strong correlation between country of hosting and IDN scripts
- Japanese, Chinese, Korean and German are the most popular languages for content associated with IDNs
- Arabic script IDNs are associated with blogs, ecommerce and online business sites in Persian and Arabic language

Universal acceptance

- Universal acceptance of IDNs the key challenge to mass uptake
- Google Gmail began supporting internationalised email addresses starting in July 2014. Popular open source email services are also supporting IDN emails
- Standardised programming tools for mobile application developers support IDNs
- Social media and search have improved support for IDNs as URLs in links
- Universal acceptance is a wider issue than previously thought. Work needs to be done on multiple fronts to ensure that IDNs can be used seamlessly

The IDN World Map 2013

Country of hosting vs IDN script



*Han includes Han, Katakana, Hiragana, where Han is the predominant script.
Japan also has high numbers of Hiragana and Katakana domains hosted.



Contents

	FOREWORD BY NEELIE KROES	10
1	EXECUTIVE SUMMARY	12
2	INTRODUCTION	18
3	Focus A Why IDNs are drivers of multilingualism	
3.1	STATUS OF MULTILINGUAL ONLINE CONTENT	20
3.2	HOW DO IDNS ENHANCE LINGUISTIC DIVERSITY ON THE NET?	23
3.2.1	IDNs help to enhance linguistic diversity in cyberspace.	24
3.2.2	The IDN market is more balanced in favour of emerging economies	25
3.2.3	IDN scripts are accurate predictors of the language of content	27
3.3	ANALYSIS OF .EU IDNS	28
3.3.1	Languages cluster around relevant scripts	28
3.3.2	.eu IDNs show strong correlation with website language	30
3.4	USAGE RATES OF IDNS	31
3.4.1	Proportion of IDNs with active name servers.	31
3.5	IDNS BY COUNTRY OF HOSTING	33
3.5.1	.eu IDNs	36
4	Focus B Reading, visual perception and language competence: a complex relationship by Giovanna Marotta and Margherita Donati	
4.1	THE COMPLEX RELATIONSHIP BETWEEN WRITTEN LANGUAGE AND VISUAL PERCEPTION	39
4.2	HOW THE BRAIN PERCEIVES WRITTEN LANGUAGE: (NEURO-)COGNITIVE STUDIES.	39
4.3	PROCESSING A WRITTEN SECOND LANGUAGE (L2)	41
4.4	CONCLUDING REMARKS	42



5 **Focus C** Universal acceptance of IDNs by Mark McFadden

5.1	WHAT DOES UNIVERSAL ACCEPTANCE MEAN?	44
5.1.1	Universal acceptance, IDNs and recent experience	45
5.1.2	Universal acceptance and new internet tools	45
5.1.3	New domain names	45
5.2	UNIVERSAL ACCEPTANCE AND MOBILE DEVICES	46
5.2.1	Is mobile different?	46
5.2.2	Built-in support.	46
5.2.3	Is the universal acceptance problem solved for mobile?	47
5.3	IDN SUPPORT IN WEB-BASED SERVICES	48
5.3.1	Overview	48
5.3.2	IDN URL support in web-based services - IDNs as content	48
5.3.3	IDN user identifiers	50
5.4	IDNS BEYOND THE USER.	50
5.4.1	Digital Certificates.	50
5.4.2	Policy data in the DNS	51
5.5	UNIVERSAL ACCEPTANCE AND EMAIL	52
5.5.1	A difficult problem to solve	53
5.5.2	Implementation status for internationalised email	53
5.5.3	Email clients – traditional clients	54
5.5.4	Email clients – web-based services	55
5.5.5	Email clients – mobile and portable devices	56
5.6	IDNS AND BROWSERS.	56
5.6.1	How a browser decides which IDNs to display	56
5.6.2	IDNs and the mobile browser	58
5.7	CONCLUSION	58



Contents

6	Focus D	
	Industry opinions	
6.1	A REMINDER OF THE DOMAIN NAME SUPPLY CHAIN	60
6.2	METHODOLOGY – REGISTRY SURVEY	61
6.3	RESULTS OF REGISTRY SURVEY	62
6.3.1	How does the uptake of IDN registrations relate to your expectations?	62
6.3.2	How well are IDNs supported by your registrars?	63
6.3.3	How would you rate end-user awareness of IDNs?	65
6.4	WHAT SINGLE CHANGE WOULD INCREASE UPTAKE OF IDNS?	66
6.5	REGISTRAR SURVEY	67
6.6	CONCLUSIONS	68
7	Focus E	
	IDNs in Arab States	
7.1	IDN READINESS MATRIX	71
7.2	FOCUS ON ARAB STATES	72
7.3	OVERVIEW – GROWTH OF INTERNET IN ARAB STATES	73
7.4	IDN ADOPTION IN ARAB STATES	73
7.4.1	Specific challenges in the region.	75
7.4.2	How are Arabic domain names being used?	76
7.4.3	Strong correlation between local language content, country of hosting and script of IDN.	78
7.4.4	A new approach in the region.	78
7.4.5	Other developments in the region	79
7.4.6	Task Force on Arabic Script Domain Names	79

8 Focus F New gTLD IDNs

8.1	OVERVIEW	82
8.2	MEETING UNMET NEEDS AND SUPPORTING MULTILINGUALISM	82
8.2.1	New gTLD applications – do they represent the world's most popular languages?	82
8.2.2	Early market performance of new gTLDs (including IDNs)	83
8.2.3	Case study on شبكة	84
8.2.4	Case study on みんな	85
8.2.5	Case study on 中文网(.Chinese) and 在线 (.Online)	86

9 Focus G IDN facts and figures

9.1	TOTAL IDN REGISTRATIONS (IDN ccTLDs, IDN gTLDs AND IDNs AT SECOND LEVEL)	88
9.2	DEPLOYMENT OF IDNs BY DOMAIN NAME REGISTRY	88
9.3	IDNs PERFORM STRONGLY IN RUSSIAN FEDERATION, ASIA AND THE PACIFIC	90
9.3.1	IDNs by region 2009-2013 (IDN ccTLDs and IDNs at second level)	92
9.3.2	IDN ccTLDs	94
9.3.3	How does IDN growth rate compare with the growth of ASCII domain names?	94
9.4	CONCLUSION	98
	APPENDIX 1: WHAT ARE INTERNATIONALISED DOMAIN NAMES, AND WHY ARE THEY IMPORTANT?	101
	APPENDIX 2: COUNTRY CASE STUDIES	106
	APPENDIX 3: METHODOLOGY FOR LANGUAGE ANALYSIS	120
	APPENDIX 4: RESULTS OF USABILITY STUDY 2014	122
	APPENDIX 5: STATUS OF IDN ccTLDs (AS AT JULY 2014)	123
	ENDNOTES	126
	GLOSSARY OF TERMS	131



Foreword

Neelie Kroes, Vice President European Commission

As the role and importance of the Internet become ever more obvious, we have been working to ensure that the Internet is open and inclusive of all stakeholders, compliant with human rights and respectful of the rule of law.


We are doing that not just for the sake of those important principles, but because an open and inclusive Internet is crucial for growth and development, for economic and social welfare. For that reason our Internet policies should help diverse cultures and communities to exploit the opportunities that the Internet presents, while preserving that diversity.

This is where the ability to access the Internet in one's own language becomes so important.

Across the world I have seen the role that new technology can play. The Internet is a platform for amazing innovation, and one that is able to cope with diversity, and adapt to local needs and sensitivities. Across the world it can cut poverty, promote and protect fundamental rights, empower individuals and groups by connecting them to unlimited opportunity.

The European Commission, like many other stakeholders, is working for the Internet to remain a single, open, free, unfragmented network of networks. But this network does not have to be homogenised, it can give support to and benefit from linguistic and cultural diversity. The underlying power of the Internet can be used to ensure that the user interface is much more sensitive about and adapted to the cultural and linguistic needs of the local user.

So I believe that internationalised domain names are an essential entry point for a multilingual and inclusive Internet, and that is why we have made them an integral part of our policies to protect an Internet that is open and accessible to all.



We must include as many people as possible in the digital opportunity that the Internet provides.

It's about communities finding their own solutions, innovating and adapting for different needs and different cultures. This is a tool to empower, not a straitjacket to impose. This is a chance for the developing world not just to develop – but to leapfrog ahead.

And as this report shows, the use of IDNs can lead to a more proportionate use of the Internet in relation to the native languages of user groups. Not by driving English speakers away, but by empowering and facilitating speakers of other languages to come online. IDNs are also a way of redressing the balance in favour of multilingual content.

Today, the Internet is a precious gift – but two thirds of the planet is still not receiving it. We need to bring those benefits and that potential to more people. We need to say no to a digital divide – yes to seamless connectivity serving every citizen.

That needs infrastructure. It needs the broadband networks – seamless, fast, and pervasive – that alone can ensure and enable this tremendous technological and social transformation. But it also brings us back to the challenge of accessibility of the Internet itself – where language can be a major barrier even if the networks and equipment are there.

Only by pursuing a policy of multilingualism in Internet access can we give effect to our policy of an Internet that is truly global, and truly accessible to all.

And that makes it a challenge worth fighting for.

Neelie Kroes
Brussels 2014

1 Executive Summary

Neelie Kroes, in the foreword to the 2014 World Report on Internationalised Domain Names, says:

“...this network does not have to be homogenised, it can give support to and benefit from linguistic and cultural diversity”

Twenty years ago, 16 million people were online. Today, the number has grown to nearly 3 billion. Now, over half the world’s Internet users are from Asia, people with different writing systems to those of the world’s first Internet users (in North America and Europe).

Domain names are an essential and ubiquitous part of the Internet. At the layers that are visible to people (browsers, email, search results, links), they provide multi-layered visual and linguistic clues about the sort of content, organisation – even the level of trust or reliability – that will be found at various online destinations. But domain names are also at work in layers of the Internet that are not visible to users, such as digital certificates, or policy data carried in domain name headers. These contribute to the seamless running of the Internet.


Traditionally, domain names have only included the Latin characters “a” to “z”, digits “0” to “9” and the hyphen “-“. Cognitive science teaches us that context and the ability to understand a language are key components in recall. It follows that those majority of Internet users unable to read or understand the names and words comprised in domain names, cannot access those clues. A multilingual domain name environment is the only way to ensure that each end user has the same rights to access content in their own language, and to experience the Internet without constraints and barriers.

Internationalised domain names (IDN) were first launched at the second level (eg παράδειγμα.eu) from 2000. From 2009, it became possible to register domain names entirely in non-Latin scripts (eg 例子.中国). At the end of 2013, there were 6 million IDNs (includes both second level and IDN TLD). Although this is a large number, it is just 2% of the world’s registered domain names (270 million).

1.1 IDNs signal local language content

The term “*internationalised* domain name” carries an inherent assumption that the default position will be ASCII, Latin script, or English language. To be other is to be “internationalised”.

The evidence shows that, far from being “internationalised”, IDNs are intensely localised. They are strongly linked to local language content, and although they occur in diverse



writing systems, the location of such scripts is closely coupled to countries and regions where related languages are spoken. An example is seen in the IDN hosting map. Cyrillic script domains are found in Russian Federation, Ukraine, Kazakhstan, Belarus and Bulgaria. Arabic script domains are found in Arab States and the Islamic Republic of Iran, and so on. Another example of localization is that ccTLDs, when deploying IDNs, tend to deploy only scripts (or even a small set of characters) that are needed to support local languages.

IDNs help to enhance linguistic diversity in cyberspace. The languages of Asia (Japanese, Chinese, Korean) are more likely to be found associated with IDNs than with traditional, ASCII domain names.

There are other ways in which IDNs promise greater diversity online. For example, the IDN market is more evenly distributed than that of general domain names, with TLDs from Viet Nam, Russian Federation, China, Taiwan of China, Japan and Republic of Korea enjoying between 10-20% market share each.

IDNs are accurate predictors of the language of web content. Our analysis of the language of web content associated with IDNs is that they show a near perfect correlation between language of web content and script of IDN. Analysis of .eu domains show that languages cluster around relevant scripts: for example, Greek language websites within the IDN sample are *only* associated with Greek script IDNs.

1.2 Universal acceptance – a difficult problem to solve

Given the strong indicators that IDNs are useful as signposts to local language content, why are there still comparatively few IDN registrations?

Part of the answer is “universal acceptance”, the ability to use IDNs across all applications and services associated with domain names. Our conclusion is that IDNs are difficult to use and do not work at all in many contexts. Our review of universal acceptance builds and expands upon on the 2012 World Report. It finds that while significant progress has been made since last year (including Google’s recent announcement that Gmail will support IDN email addresses from 2014), much remains to be done. Most email providers (traditional, webmail and mobile) do not support internationalised email. Many online services require email addresses as the user identifier to set up online accounts – none support internationalised email addresses in this context. Universal acceptance means more than just browsers and apps. This year’s study includes a review of universal acceptance of IDNs beyond the user: in digital certificates and policy data.

However, universal acceptance does not just affect IDNs. Similar issues affect all new TLDs, and it may be that the launch of more than 1 000 new gTLDs as a result of the ICANN programme is providing an impetus for providers to deploy resources to ensure that all new TLDs (including IDNs) work seamlessly at all levels of the Internet stack (infrastructure, addressing and applications). Major vendors acknowledge that, because

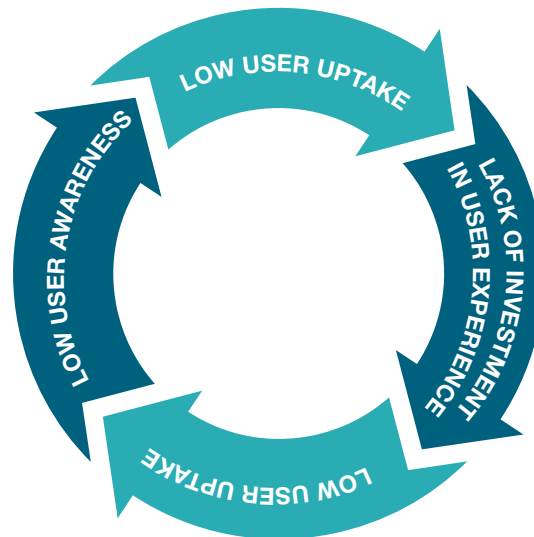
of the ubiquity of domain names across all applications and services, ensuring universal acceptance is a considerable task.

1.3 IDNs and industry opinions

The lack of universal acceptance, and the consequent lack of visibility of IDNs to consumers, may be influencing industry opinions. Each year, we conduct a survey of registries and registrars active in the IDN space. This year, thanks to the collaboration of the regional ccTLD organisations (APTLD and LACTLD in addition to CENTR), we received more responses to our registry survey than ever before. Across all measures, it appears that confidence is lower than previous years. The single change that most registries would advocate is universal acceptance of IDNs across all applications and services.

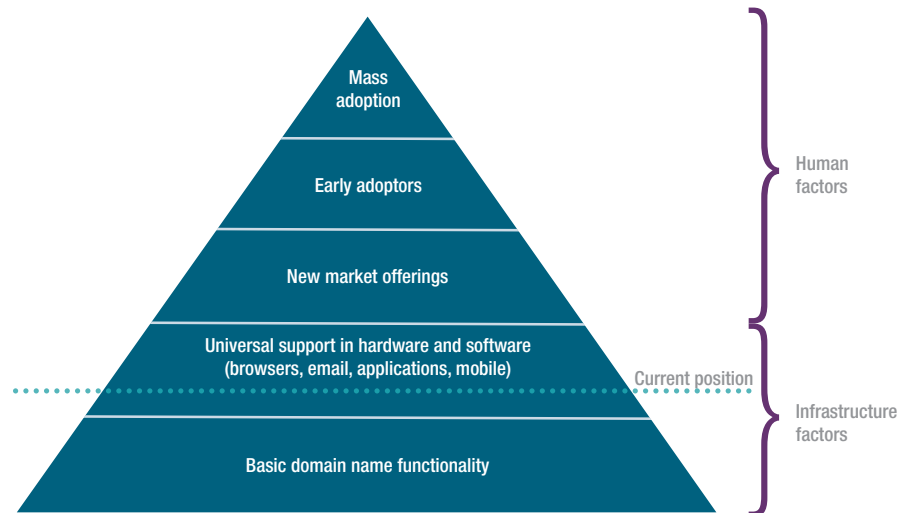
The wider environment is currently creating a negative cycle of poor user experience, low user uptake, and low user awareness, which itself leads to low user uptake, and so on (see figure 1).

Figure 1 – IDN negative cycle



To break the negative cycle, progress is needed in both universal acceptance and user awareness. Universal acceptance is a vital prerequisite for mass uptake, as indicated in our “IDN hierarchy of needs” (figure 2), adapted from Maslow’s well-known hierarchy of needs model.

Figure 2 – IDNs - Hierarchy of needs




On this view, the deployment of IDNs is still part way through the “infrastructure factors” (see figure 2), the lowest levels which form the foundation for future growth, and human use. Without basic functionality, and support across hardware and software, sustained mass uptake will prove elusive. Although the diagram in figure 2 suggests a steady progression from one step to the next, in reality, IDNs are already on the market, and have been registered by early adopters.

1.4 IDNs in Arab States

There are 125 million Internet users across the Arab States and Islamic Republic of Iran. Social media, particularly Twitter and Facebook, are popular in the region, and Arabic is growing as a preferred language of user generated content.

Despite burgeoning local language content in Arabic, registrations of domain names (whether ASCII or IDN) remain low across the region, even in countries such as Egypt where many factors point to a high potential for mass uptake of IDNs. Colleagues in the Egyptian registry highlight specific challenges in the region, including an immature domain name market and related industries, lack of user awareness, complex registry policies, lack of universal acceptance and lack of appropriate skills and experience in the region (capacity).

Nevertheless, IDNs (where they are in use) show familiar correlations with hosting country and local language content. Content in Arabic or Persian tends to be “high involvement” (ie requiring effort by the content creator) such as blog, ecommerce and business sites.



There are some signs of change. Some have liberalised their policies with successful results. Registries in the region (as elsewhere) are tireless advocates for IDNs, and have been conducting research on universal acceptance and models of working IDN email systems. Experts from the region are working through ICANN to develop rules for handling Arabic script variants across multiple languages – an important contribution to universal acceptance.

1.5 ICANN's new gTLD programme and IDNs

2013 brought the first of ICANN's new gTLDs to market, including a handful of IDN new gTLDs. Despite its objective of meeting unmet needs in the domain name system, the new gTLD programme has not succeeded in introducing greater linguistic diversity. Only 6% of applications are not in Latin script, and analysis of the language of the individual strings (eg .photography etc) shows that 90% are either in English language or understandable to English speakers.

Early market performance of new gTLDs shows a familiar 'long tail' pattern, and the IDNs conform to that pattern. Case studies of three IDN new gTLDs show mixed performance, even at this early stage.

1.6 Deployment of IDNs

We are now aware of 74 IDN deployments, 50 at the second level, and 24 IDN TLDs. A review of IDN launches since 2000 show that whereas 10 years ago, registries were launching at the second level, today the growth in number of launches is greater in IDN TLDs.

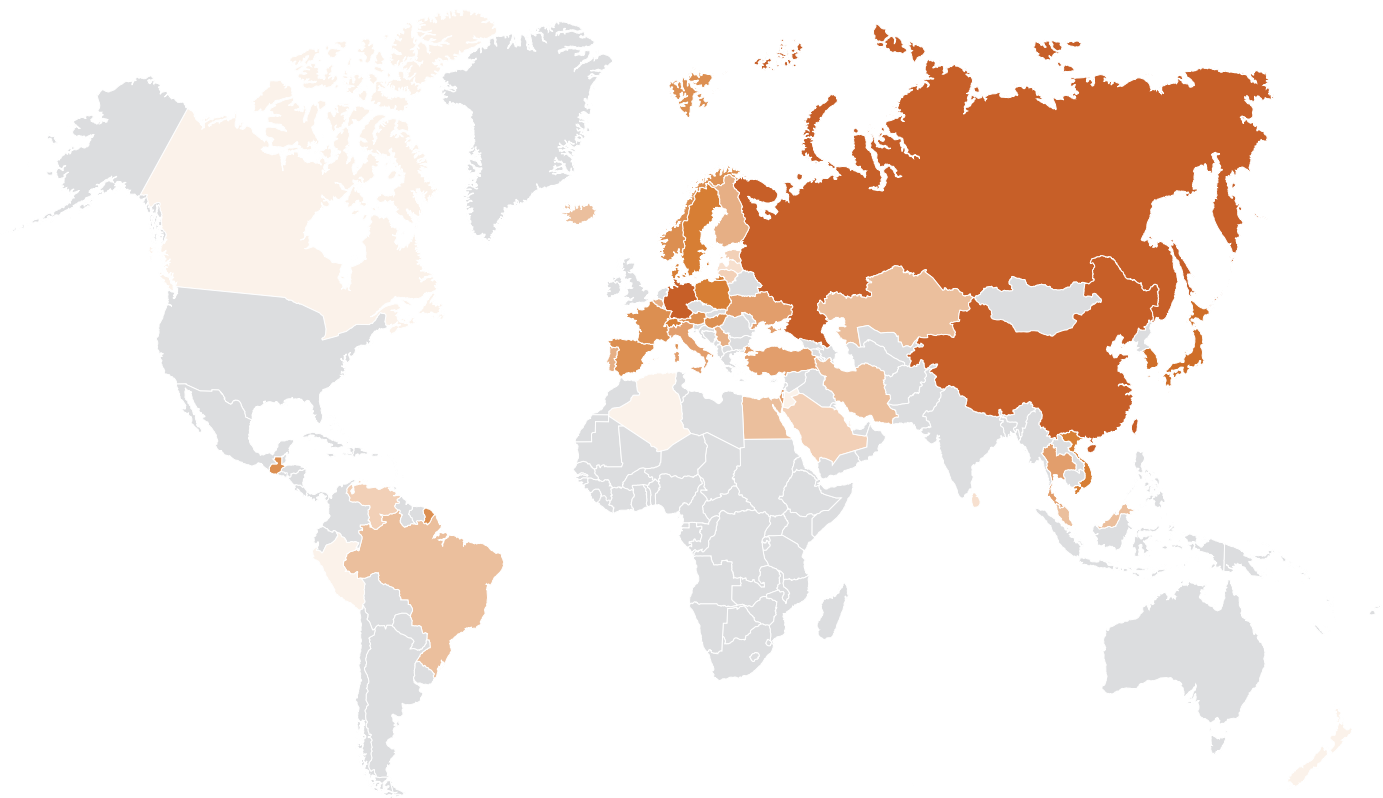
Continuing the theme of localisation, we see that IDNs perform strongly in Russian Federation, Asia and the Pacific. Although there are lower numbers of IDN registrations across the Arab States and Latin America, growth is positive, and new deployments of IDNs continue each year.

Taken as a whole, IDNs tend to be growing at a more rapid rate than traditional, ASCII domain names (however, numbers are lower, so higher percentage growth is more readily achievable). IDN TLDs tend to have greater volatility of growth patterns at this time.

IDNs have great potential to enhance linguistic diversity, essential for an inclusive Internet environment. Without universal acceptance, that potential will not be fulfilled. To make progress requires significant effort by a diverse group of actors throughout the value chain and Internet stack. It is hoped that significant progress will be made in this area in the coming years, so that everyone can enjoy the benefits of the Internet, without constraint and barriers.



Figure 3 – The IDN World Map 2013 –
Excludes gTLDs. Includes ccTLDs (both full IDN, and hybrid IDN)



> 500 000	50 000 - 99 999	10 000 - 19 999	2 000 - 4 999	500 - 999
100 000 - 499 999	20 000 - 49 999	5 000 - 9 999	1 000 - 1 999	1 - 499



2 Introduction

This report begins with an analysis of why IDNs are drivers of multilingualism (Focus A). It looks at the ways in which IDNs enhance linguistic diversity in cyberspace. It then reviews .eu IDNs by language of content, usage rates, and country of hosting.

Focus B considers the way that the brain processes language, in a chapter by Giovanna Marotta and Margherita Donati of the University of Pisa. This section describes the complex relationship between written language and visual perception.

Focus C provides an in depth analysis into universal acceptance of IDNs, in a chapter by Mark McFadden, following on from last year's report. The section reviews progress over the past 12 months, and highlights issues relating to mobile devices, web-based services, IDNs beyond the user (in digital certificates and DNS policy data), IDN email, and IDNs in browsers.

There follows a report on our annual registry and registrar survey, which picks up themes from the previous sections.

Focus D returns to the IDN experience in Arab States. It reviews our IDN readiness matrix, and provides an overview of IDN growth across the region. It considers specific challenges in the region, how Arabic domain names are being used and considers correlation between local language content, country of hosting and IDN script. It rounds up by looking at individual registry experiences and other developments in the region, and the work of the Task Force on Arabic IDNs.

2013 saw the launch of the first of over 1 000 new gTLDs. Focus E provides an overview of the process, analyses IDNs by language, looks at the early market performance, and provides brief case studies on three IDN new gTLDs.

Finally, the report reviews adoption of IDNs over the past 12 months, including total IDN registrations, deployment of IDNs by registry and a review of IDNs by region. It considers growth rates of IDNs compared with general (ASCII) domain names.

The appendices include country case studies for 9 countries across three geographic regions: Asia and the Pacific; Arab States, and Europe.

FOCUS A

Why IDNs are drivers
of multilingualism



3 Why IDNs are drivers of multilingualism

3.1 Status of multilingual online content

There are 7 billion people in the world, and approximately 6 000-8 000 languages in use¹. There is not a 1:1 correlation between nationality and languages spoken, however. Throughout the world, “monolingualism is not the rule, but the exception”². Linguistic diversity is particularly high in Central African Republic, Papua New Guinea, Belgium, Belize, Vanuatu³, and many populous countries, such as South Africa, India and Nigeria, have high linguistic diversity.

To what extent is the offline linguistic diversity reflected online? The situation is improving, but English is still the dominant online language. In 2014, English language represents over 55% of web content⁴, compared with 75% in the late 1990s⁵.

There is international consensus on the need to promote linguistic diversity, in cyberspace as well as offline, as reflected in the World Summit on the Information Society (WSIS) action line C8 (Cultural diversity and identity, linguistic diversity and local content) and UNESCO’s *Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace* (2003).

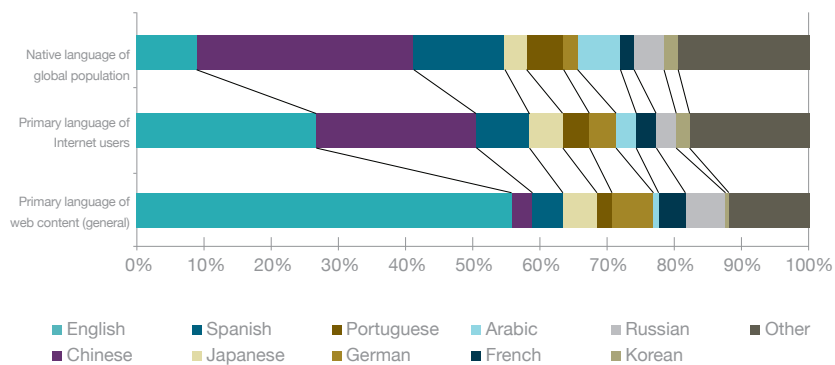
In part, the dominance of English language content is a legacy from the Internet’s US origins. It may also be a consequence of English being a popular second language⁶. A study by the Internet Society⁷ indicates that the the primary language of Internet users does not yet balance with that of the global population, and the primary language of the world’s 10 million most popular websites further accentuates this imbalance (see figure 4).

Some of the world’s most spoken languages offline, such as Hindi, Bengali, and Javanese⁸, do not currently feature in the primary language of Internet users, or primary language of web content.

It is a challenge to find appropriate indicators for measuring linguistic diversity online⁹. As in many other areas of life, we tend to measure what is measurable and visible – for example, web content¹⁰. However, multilingualism in cyberspace also involves programming languages and environments, email, messaging, chat, blogs and content on social networks. For example, the Arab Social Media Report¹¹ shows that Arabic is the primary language of tweets and Facebook posts for users within Arab States.

The popularity of English language content online is disproportionate compared with offline life

Figure 4 – Proportion of web content associated with IDN domains and ASCII domains, for the top ten internet user languages¹²



Reflecting the increasing linguistic diversity amongst the user based for popular web services, support for multilingualism is growing. Table 1 updates measures published in the Broadband Commission Report 2012.

Table 1 - Support for linguistic diversity by popular web services

Name of service	Number of users	Languages supported ¹³	Notes
Twitter	255 million active monthly	35+ ¹⁴	Network of 350 000 translators work through Twitter translation centre ¹⁵
Google Translate		80	Statistical machine translation – based on patterns in large amounts of text, users are encourage users to contribute improved translations ¹⁶
Facebook	1.3 billion active monthly	73 ¹⁷	Facebook also relies on a network of users who contribute translations ¹⁸
Wikipedia	21 million ¹⁹	287	Number reflects languages for which official Wikipedias have been created ²⁰ . 9 languages have over 1 million Wikipedia articles.

Popular applications are supporting multilingualism in cyberspace

The number of languages supported by major web applications has expanded in the last 5 years. While the numbers are still far short of the estimated 6 000-8 000 still in use in the world, the distribution of language speakers is not evenly spread. The top 74 languages are spoken by 94% of the world’s population²¹, so Google and Facebook’s environments are close to that number, and Wikipedia supports many minority languages.

Box 1 – Social networks can support minority languages

A study of the use of Facebook by the Eton of Cameroon, a language with 250 000 speakers, captures the benefits of multilingualism in cyberspace. It found that forms of writing in the language were evolving within Facebook groups, and that “among the many topics covered in these Facebook groups, a significant portion is devoted to generic issues relating to “traditional” culture, marriage, parenthood, initiations and sayings”.

Net.Lang, 2012, “The use of Facebook by the Eton of Cameroon”, Rivron, V., p 161 ff


A comparison between languages with over 1 million Wikipedia articles, and domain name registrations per 1 000 of population produces some interesting correlations (see table 2).

The table indicates a correlation between uptake of domain names at the ccTLD level (particularly within Europe) and creation of Wikipedia articles. Many European countries have a high level of domain name penetration (measured by 1 000 of population). The Oxford Internet Institute has shown that in 2013, more Wikipedia articles were created by users within Europe than the rest of the world put together²².

Table 2 – Correlation between number of wiki articles and domain name penetration by country

Country /ccTLD	Language	Number of Wiki articles (in millions)	Number of domains in ccTLD (in millions) ²³	Domains per 1 000 population ²⁴
United Kingdom, .uk	English	4.5	10.5	163
Netherlands, .nl	Dutch	1.8	5.4	305
Germany, .de	German	1.7	15.6	190
Sweden, .se	Swedish	1.7	1.3	131
France, .fr	French	1.5	2.7	38
Italy, .it	Italian	1.1	2.6	42
Russian Federation, .ru, РФ	Russian	1.1	5.7 (.ru+ РФ)	30
Spain, .es	Spanish	1.1	1.7	34
Viet Nam, .vn	Vietnamese	1.1	1.2	13
Philippines, .ph	Waray-Waray	1.1	[no data]	[no data]

Clearly, many of the languages listed (eg English, German, Russian, Spanish) are spoken in numerous countries, and therefore the comparison is incomplete. It also misses gTLDs (eg .com and .net) or regional domains such as .eu which are registered across many countries and territories. The table highlights the ccTLD with the largest number of domains.



The context in which internationalised domain names exist is therefore one of burgeoning multilingualism in cyberspace (much of it user-generated), and significant commitment to linguistic diversity by both international organisations and major service providers. Moreover, localised services have hundreds of millions of users, for example Yandex (search) in Russian Federation²⁵; Nate On (social network) in Republic of Korea²⁶; Tencent QQ (microblogging) in China²⁷.

3.2 How do IDNs enhance linguistic diversity on the Net?

The domain name system is key to finding content on the Internet. Although many users rely on search and apps, domain names continue play an essential and ubiquitous role throughout the Internet stack (infrastructure, addressing, applications).

In last year's study, we showed that IDNs are a predictable signal of the language of online content. Analysis of the language of websites associated with .eu, .com and .net IDNs showed a near perfect correlation between the language of web content, and script of IDN (at the second level, eg παράδειγμα.eu). So, Cyrillic IDNs were 100% likely to point to Russian, and Bulgarian language web content; Hangul IDNs were 100% likely to point to Korean language web content; Han to Chinese and Japanese; Arabic to Persian and Arabic, and so on.

This year we have repeated our cooperation with Verisign, and at the same time have extended our research into correlation between IDNs and language of website. In addition to the .com and .net IDNs analysed by Verisign, we have also analysed all 51 000 .eu IDNs (rather than a representative sample of 10 000 last year), and IDNs in open zone files (gTLDs). This has given us a dataset of over 1.6 million IDNs. The methodology is described in appendix 3.

As a result of our analysis, we are confident in making the following three statements:

- ***IDNs help to enhance linguistic diversity in cyberspace***
- ***The IDN market is more balanced in favour of emerging economies***
- ***IDNs are accurate predictors of the language of web content***

3.2.1 IDNs help to enhance linguistic diversity in cyberspace

We have seen above (section 3.1) that English is over-represented online compared with the number of speakers offline, and the primary language of Internet users²⁸. Latin script languages which use accents and diacritics (eg Spanish, Portuguese) are reasonably well represented. Meanwhile, languages associated with non-Latin scripts which are widely spoken offline (eg Chinese, Korean, Arabic) are under-represented compared with the primary language of Internet users.

When the analysis is expanded to include language of web content associated with IDNs, a different pattern emerges (see figure 5, lowest bar). English is the language of 10% of IDN websites, compared with 55% in general websites. This is closer to the proportion of native speakers in the global population. The languages of Asia (Japan 40%, Republic of Korea 10%) and emerging economies (Chinese 10%) are better represented as the language of web content associated with IDNs, than in general websites. There are some anomalies, for example the strength of Japanese language content associated with IDNs is out of proportion of the primary language of web users. Meanwhile Arabic, and Russian content are better represented in general web content than they are in content associated with IDNs.

There could be a number of reasons for this. First, our data sample includes gTLDs and .eu as the only country code, but not other country code TLDs. It is likely, for example, that the inclusion of 800 000 Cyrillic script IDNs registered under рр would affect the balance of Russian language websites; similar for 750 000 Han script IDNs under 中国/中國, and 台湾/台灣.

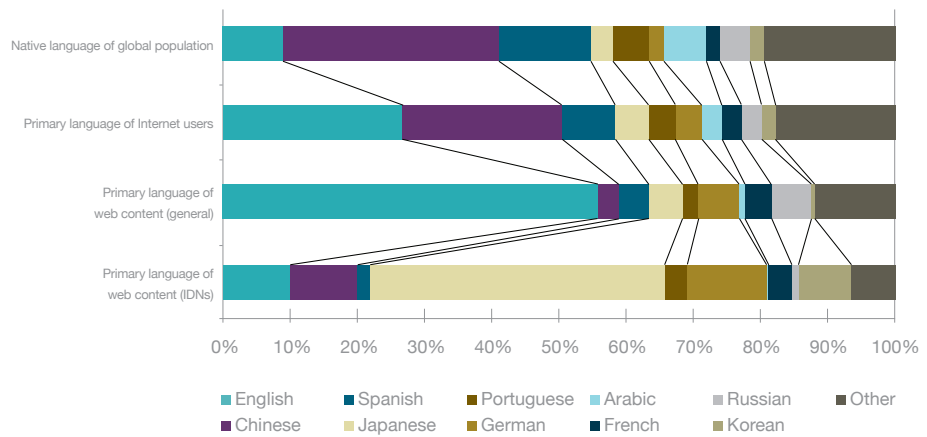
We know from our study of domain name registrations in the Arab States that overall numbers are still low, and this pattern is seen also in our data sample, where Arabic script IDNs formed only 0.4% of the total.

Our analysis, and that of Verisign, relied on automated translation tools (such as Google Translate) for bulk analysis of websites. The automated tools are usually accurate, but occasionally produce anomalous results. For example, we found they sometimes incorrectly identified Portuguese as the language of web content instead of Japanese, and incorrectly identified Greek language instead of Korean. Otherwise, manual checking confirms a high level of confidence in the other results.

We can therefore say with confidence that IDNs enhance online linguistic diversity. They are strongly associated with languages of Asia and emerging economies which use non-Latin scripts. Not only can IDNs help to detect or predict existence of content in multiple languages, but they can also be seen as a means of accessing content in multiple languages.



Figure 5 – Proportion of web content associated with IDN domains and ASCII domains, for the top ten internet user languages



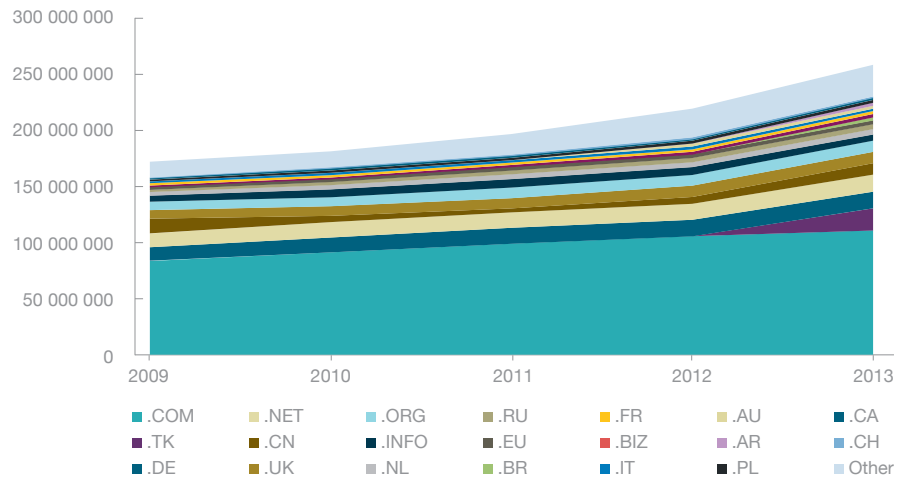
IDNs signal multilingual content, and help redress the imbalance of online languages

3.2.2 The IDN market is more balanced in favour of emerging economies

We have analysed market share of general domain names and IDNs within the period 2009-2013.

The strength of .com’s market share is a feature of the general domain name market (see figure 6), (48% (2009) reducing to 42% (2013)). Even the most successful of the

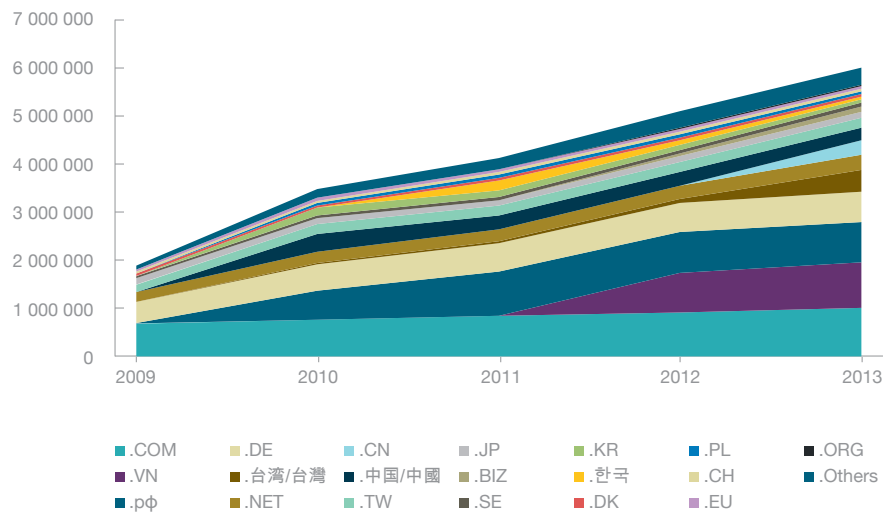
Figure 6 – Domain name market share over time (all TLDs)



.com has 42% market share of total domain names



Figure 7 – Top twenty IDN registries: market share over time



The IDN market is more balanced in favour of emerging economies

other TLDs have never achieved over 10% market share in the period. European and North American TLDs tend to dominate the general market out of proportion with their local populations²⁹.

The first striking thing about the IDN market (figure 7) is how rapidly it has grown in the past 5 years, from under 2 million in 2009 to 6 million in 2013 (220% growth compared with 42% growth of the general market in the same period).

Secondly, the IDN market is more evenly distributed than that of general domains. While .com is the market leader with 17% (2013), there is generally a more even distribution of market share, with six TLDs, Viet Nam, Russian Federation, Taiwan of China, China, Japan, and Republic of Korea, having between 10-20% market share.

The sample sizes are very different, and this may lead to distortions. The IDN market, at 6 million, is only 2% of the general domain name market (270 million). In numerical terms, a small change in the IDN market can lead to a large percentage difference. For example, in 2011 the 한국 TLD (Republic of Korea) had 5% of the market share. By 2013, a 150 000 reduction in 한국 within a rapidly growing market had reduced this market share to just 1%.

Overall, the IDN market is more balanced in favour of emerging economies, whose large populations (both online and offline) use non-Latin scripts.



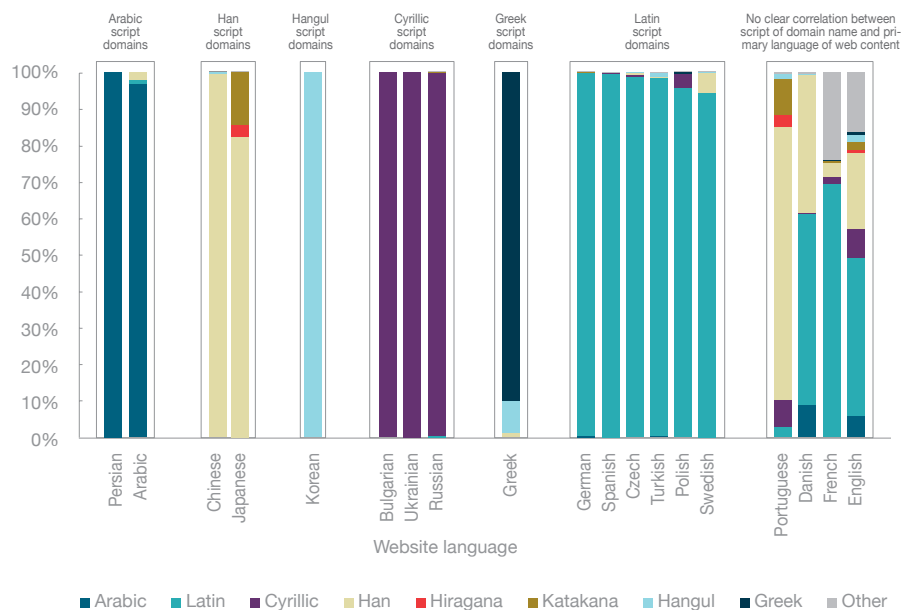
3.2.3 IDN scripts are accurate predictors of the language of content

We reviewed the language of web content associated with IDNs, to see whether there was any correlation with the script of domain name. If there were no connection between domain name script/language and the language of content, one would expect a random pattern. Alternatively, if there was a strong correlation, one might expect that a Cyrillic script domain would lead to web content in Russian, Bulgarian or Ukrainian, or that an Arabic script domain would lead to web content in Arabic or Persian, Han script to Chinese and so on.

Our analysis found that the relationship between language of web content and IDN script is not random. As we found last year, there is near perfect correlation between language of web content and the script of IDN associated with it (figure 8). In other words, IDNs are accurate predictors of what language will be found on their web content. Only English and French – which are commonly spoken as second languages around the world – are associated with a large number of scripts (Latin, Arabic, Cyrillic, Han, Katakana, Hiragana, Hangul, Greek and others), and display the more random pattern predicted in the “no connection” hypothesis.

The results for Greek, Portuguese and Danish language are overstated – as the automated translation tools wrongly identified Portuguese for Japanese, Greek for Korean language. Re-checking of the Hangul script IDNs in the data sample found no instances of Greek language websites.

Figure 8 – Correlation between website language and domain name script



Language of content and script of domain name are near perfectly related

3.3 Analysis of .eu IDNs

EURid supports the 24 official languages of the European Union. The majority of European languages rely on Latin script with diacritics and accents. Some European languages, such as German, Swedish, French, Czech and Polish, use many diacritics. Others (such as Dutch) use relatively few. Two EU languages – Bulgarian and Greek – rely on non-Latin scripts (Cyrillic and Greek respectively).

Using the same methodology as for the larger dataset (see section 3.2 above), we also analysed the language of web content associated with all 51 000 .eu IDNs. 25 000 had too little content to analyse, leaving working data set of 26 000 names. Of these, 25 000 were Latin script, 800 Cyrillic script and 200 Greek script.

If we are correct in thinking that IDNs link strongly with associated languages, we would expect to see a high correlation between script and language (eg Greek content with Greek script domain names) and to see web content in languages for which IDNs are particularly relevant (eg German, Swedish). Because the .eu domain is associated with the European Union, and has a residency requirement, we would not expect to see many non-European languages (eg Chinese, Korean) featuring in the language analysis.

3.3.1 Languages cluster around relevant scripts

As with the larger data set (section 3.2), clear patterns emerge within the .eu data.

Bulgarian and Russian language websites are associated with Cyrillic script domains (figure 9), and not with Greek or Latin script domains; Greek language websites are only associated with Greek script domains (figure 10). An array of European languages are associated with Latin script IDNs, with German language making up 57% of websites (figure 11).

The small sample sizes for Cyrillic and Greek mean that relatively small differences in numbers can result in large percentages. For example, of the 15% “other” languages in the Greek script IDNs (see figure 10), none have more than 9 websites.

As with the larger data set, English performs strongly across all three scripts reflecting its popularity as a second language amongst Internet users. French and German also appear in web content associated with a small number of Cyrillic and Greek script IDNs (less than 30).

Figure 9 – Language of websites associated with Cyrillic script .eu IDNs

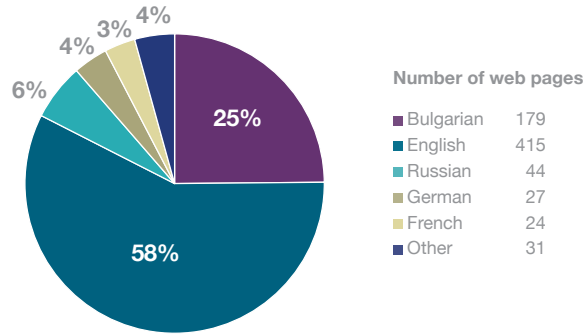


Figure 10 – Language of websites associated with Greek script .eu IDNs

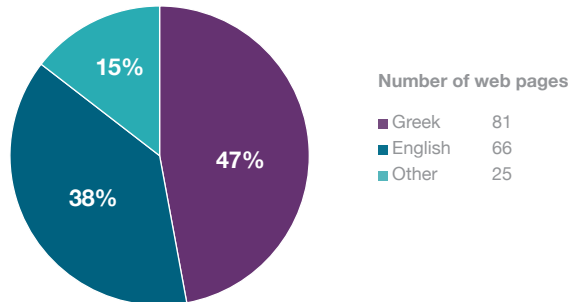
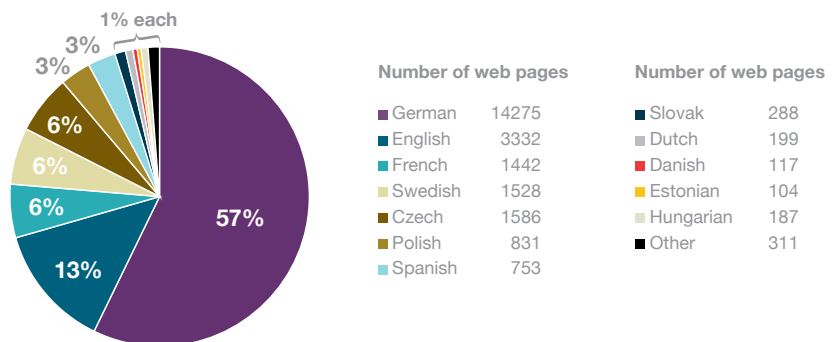


Figure 11 – Language of websites associated with Latin script .eu IDNs



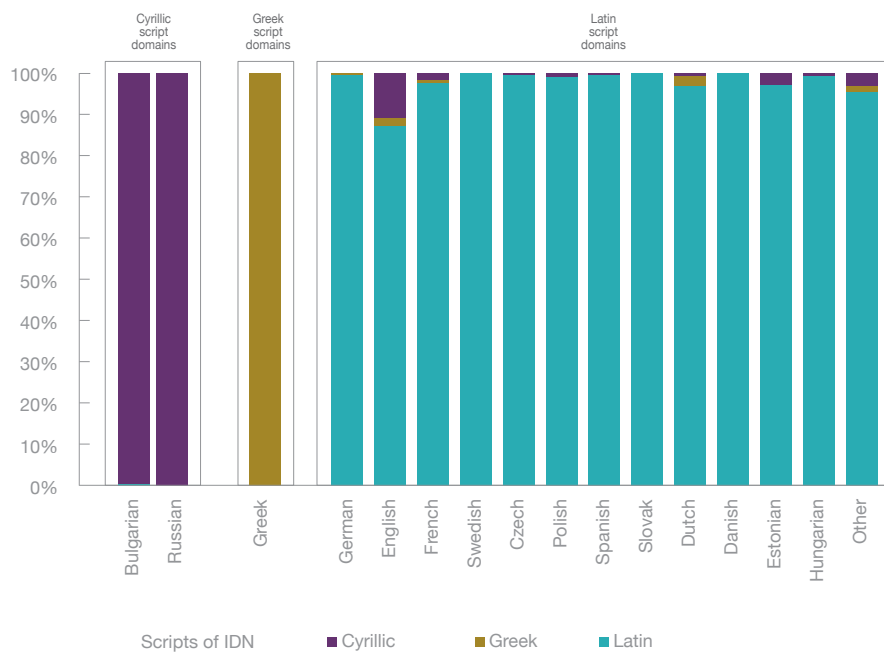
3.3.2 .eu IDNs show strong correlation with website language

Another perspective is to start with the language of web content, and analyse the script of the IDNs pointing to it.

As before, there is an almost perfect correlation between the language of web content, and the script of IDNs – it is almost always the script one would expect for the individual language. The clearest examples (see figure 12) are Greek language websites which are only associated with Greek script IDNs; Bulgarian and Russian language websites are only associated with Cyrillic script .eu IDNs (apart from a single Bulgarian language/Latin script IDN example). Only with English does the correlation dip below 90%, for reasons we have explored above.

.eu IDNs show strong correlations between IDN script and language of web content

Figure 12 – .eu IDNs: correlation between script of domain name and language of web content

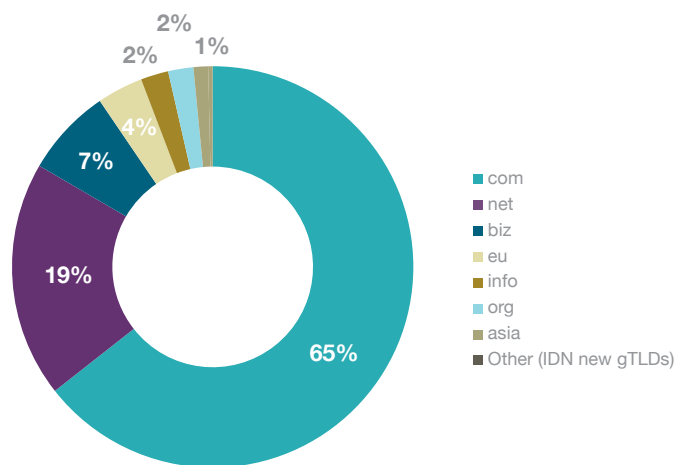


3.4 Usage rates of IDNs

This year, we examined the country of hosting for a sample of 1.9 million IDNs, including .eu, .com and .net IDNs (with the support of Verisign); and the IDNs in all open zone files, such as .org, .biz, .info, .asia and the IDN new gTLDs.

Not every registered domain name has IP addresses assigned. Our data sample comprised a total of 1.1 million domains (see figure 13). It includes both second level IDNs (eg .eu, .com, .net etc), and full IDNs (eg the new IDN gTLDs such as شبكة , みんな, 中文网 and 在线). Second level IDNs far exceed the number of full IDNs in our data sample, and this may lead to different results than a more fully balanced data sample. For this reason, we have also included pcp in our usage analysis, based on data published by the Russian registry³⁰.

Figure 13 – IDNs by country of hosting – composition of data sample



3.4.1 Proportion of IDNs with active name servers

A domain name requires active servers in order to work. We have reviewed the percentage of IDNs with active name servers.

The proportion of IDNs with active IP addresses are within a healthy range (68%-86%, with .eu IDNs second highest in the sample at 82%) (see figure 14). The rate of IDNs with active name servers is increasing year on year. For example, pcp has shown a 13% increase in active IP addresses from 2011 to 2013 (figure 15).



Figure 14 – Percentage of IDNs with active IP addresses

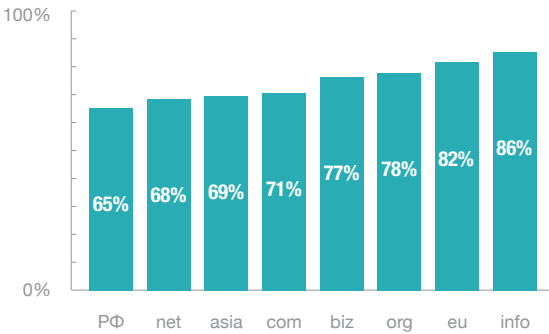
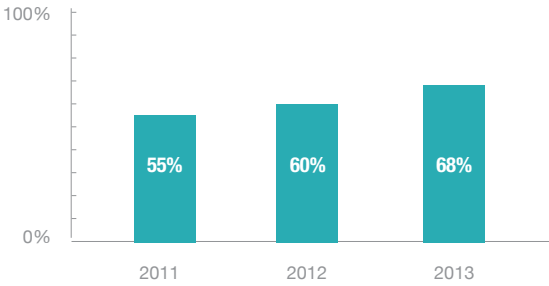


Figure 15 – Percentage of PΦ domains with active IP addresses 2011-2013

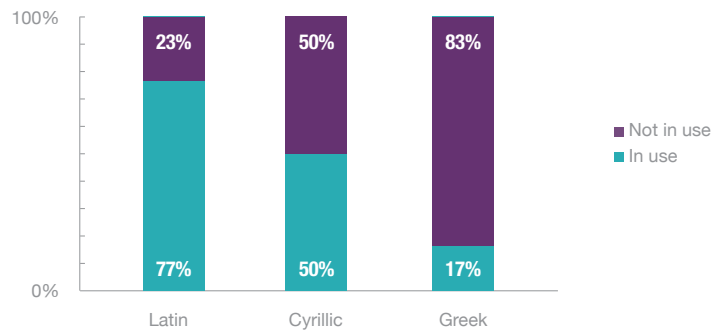


While not every domain with active IP addresses will have an active website, designating name servers is a necessary precursor to using an IDN. So, within pφ, as the rate of active name servers has increased year on year, so has the percentage of IDNs in use, from 48% last year to 50% this year. In contrast, the usage rate of .com and .net IDNs has fallen from 39% last year to 35% this year.

The .eu IDN usage rate appears to have undergone a huge increase since last year (32% to 63%). Having reviewed the data, we believe that the difference is a consequence of a change in our methodology for measuring, rather than a change of behaviour by .eu IDN users. Last year, we analysed a sample of 10 000 .eu IDNs, comprising all Cyrillic and Greek script domains, and a sample of 5 000 Latin script IDNs. Therefore the sample gave disproportionate weighting to the Cyrillic and Greek script .eu IDNs.

This year we have analysed the entire dataset of 51 000 .eu IDNs (figure 16). We have found that the usage rates amongst Latin script .eu IDNs is far higher than Cyrillic and Greek script IDNs (which are hybrid and therefore require a change in keyboard when typing). Whereas 77% of Latin script .eu IDNs are in use, only 17% of Greek script .eu IDNs are in use (ie have active web content). A similar pattern is observed within .com and .net IDNs, where 67% of Latin script IDNs are in use, whereas only 42% of Han script IDNs have active content. The well-known difficulties of hybrid IDNs help to explain these disparities in usage rates, and emphasise the importance of having fully internationalised domain names.

Figure 16 – .eu IDNs usage rates by script

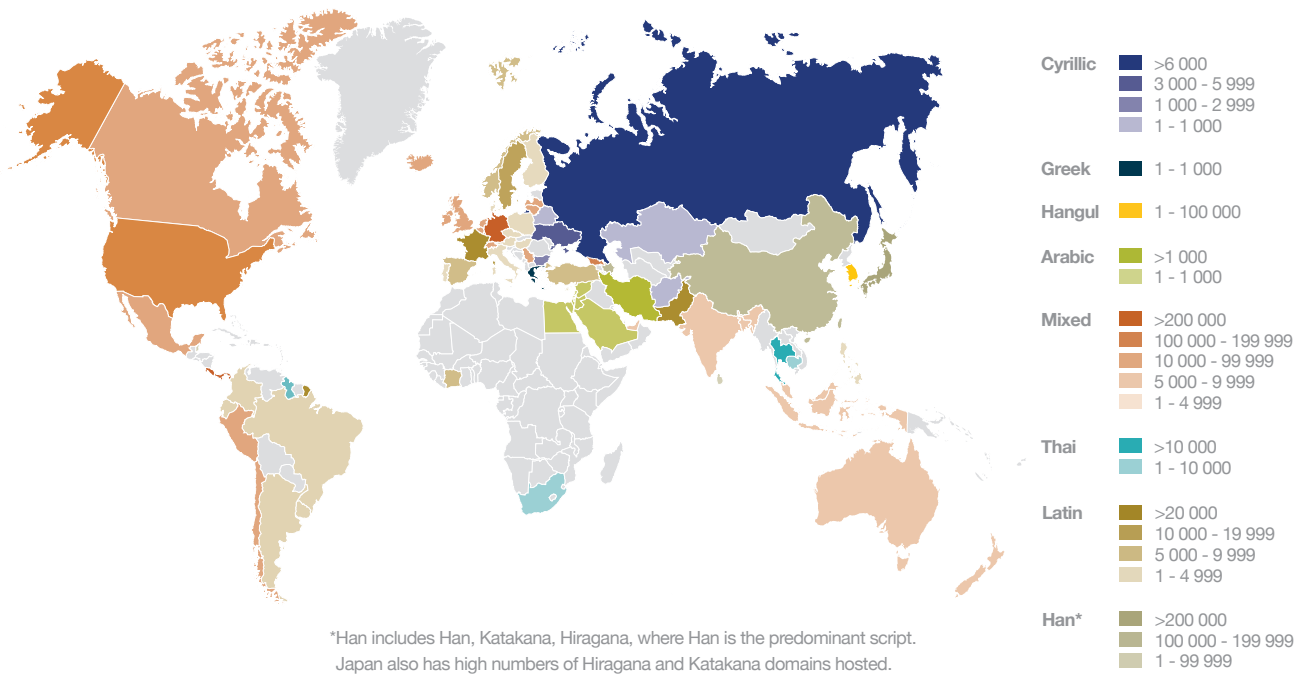


This should be encouraging news for operators of new gTLD IDNs, which started to launch in 2014. We have reviewed the percentages with active name servers for 4 IDN new gTLDs with more than 1 000 registrations (在线, 中文网, شبكة, みんな). These are currently showing low rates of active name servers (as low as 3% for the newer TLDs). However, the rates of active name servers are increasing rapidly, for example 65% active name servers in شبكة (May 2014). As can be expected in a new namespace, the percentage with active websites is still very low, but the increase in active name servers can be viewed as a harbinger of active use.

3.5 IDNs by country of hosting

Research by UNESCO, OECD, and ISOC (2011) has found a significant correlation between local servers and local language content, and that language is a local factor³¹. Our research in 2012 suggested that local registrar networks play a role in ensuring good deployment of IDNs in countries where a particular language / script is relevant. Therefore, if IDNs are acting as enablers of local language content, and bearing in mind the link between local servers and local content, we would expect to see a clusters of IDNs being hosted in countries associated with particular scripts or characters.

Figure 17 – The IDN World Map 2013 – Country of Hosting vs IDN script



Building upon the analysis of country of hosting last year, we analysed the country of hosting for 1.1 million IDNs (across .com, .net, .eu and other gTLDs). The data sample comprised a larger number of scripts than the .eu IDNs which support the official EU languages only, reflecting the larger number of scripts in the target market for gTLDs. The analysis looked not only at the country of hosting, and script of IDN. The results show that hosting patterns tend to reflect the scripts of local languages (figure 17). Table 3 shows the top five countries of hosting for IDNs in the data sample.

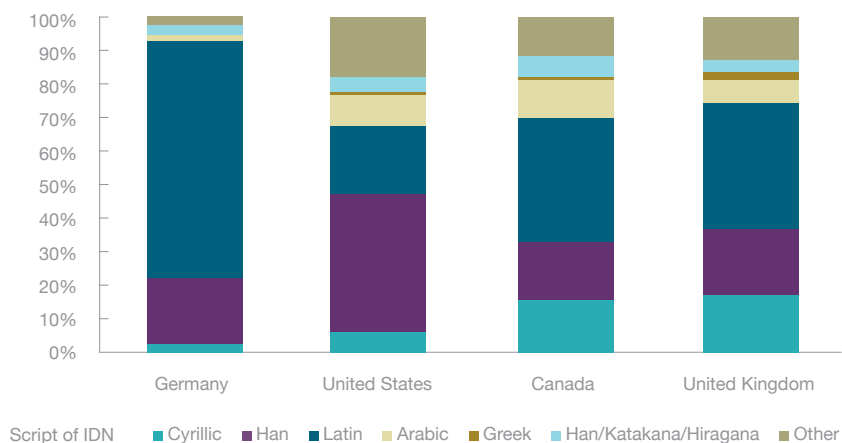
Table 3 – Top 5 countries for hosting IDNs

Country	Number of IDNs hosted
Japan	290 790
Germany	236 785
USA	141 365
China	118 815
Republic of Korea	81 575

To some degree, these results reflect the dynamics of the global registrar market where a few multinational companies based in United States, Canada, United Kingdom and Germany control a high proportion of the market.

The IDNs hosted in the four largest “mixed script” countries reflect a broad spread of scripts, especially in the English speaking countries United Kingdom, United States, and Canada (figure 18). In contrast, 71% of IDNs hosted in Germany, where the local language uses Latin-script diacritics, are identified as Latin Script.

Figure 18 – IDN script by country of hosting – mixed script countries




Over 70% of IDNs hosted in Germany are Latin script

Moving away from these "mixed script" countries, figure 18 shows how closely IDN script associated with local languages are mirrored in hosting patterns, suggesting a striking correlation between IDN script, language and country of hosting. For this analysis, we have excluded IDNs in the sample where the automated tool was unable to identify the language of the website. The percentage figure is derived from the total number of IDNs hosted in the relevant country.

Table 4 – Correlation between language, IDN script, and country of hosting

Country of hosting	Script	Primary language spoken in country	Number of speakers in country ³²	Percentage of IDNs hosted in stated country
Republic of Korea	Hangul	Korean	77.2 m	98%
Japan	Han, Katakana, Hiragana	Japanese	112 m	100%
Russian Federation	Cyrillic	Russian	168 m	98%
China	Han	Chinese	1.2 billion	100%



The visualisation (figure 17) also shows how few of the domain names in our sample are hosted within Arab States (for example, 9 in Egypt; 8 in Saudi Arabia). There are a few anomalies, but these tend to arise where there are very low numbers of hosted IDNs (eg South Africa, 1 Thai script IDN, Guyana 2 Thai script IDNs, Ivory Coast 1 Katakana IDN).

The results of our analysis of IDNs by country of hosting support research findings that correlate local servers with local content, and emphasises the link between IDNs and content in the language indicated by the IDN script.

3.5.1 .eu IDNs

At December 2013, EURid had 752 accredited registrars, representing a geographically diverse registrar-base, across the EU and overseas. Looking at all 3.7 million .eu registrations, 37% are managed by German registrars³³, and the top 100 .eu registrars manage about 84%.

If the correlation between local language content and local servers is borne out, we would expect to see clusters of hosting for .eu IDNs in countries associated with particular scripts (eg Greece for Greek script; Bulgaria for Cyrillic), and to see different patterns associated with IDN .eu domain hosting, compared with total .eu registrations. Overall, the picture may be a little blurred, as many registrars operate through networks of resellers, who may be located in any country.

Of the 51 354 .eu IDNs, it was impossible to identify the hosting country for 9 140 IDNs as no IP addresses are associated with those domains, leaving a data sample of 42 116.

A comparison of ASCII and IDN .eu registrations by country of hosting (figures 19 and 20) emphasises links between script and local language. Germany's relative share increases from 35% (all .eu domains) to 59% (IDNs), reflecting perhaps that German language uses diacritics and special characters. Of the IDNs hosted in Germany, 98% are Latin script.

Of the .eu IDNs hosted in Bulgaria, 99% are Cyrillic script; of the IDNs hosted in Greece, all are Greek script.

Figure 19 – .eu domains (all) by country of registrar

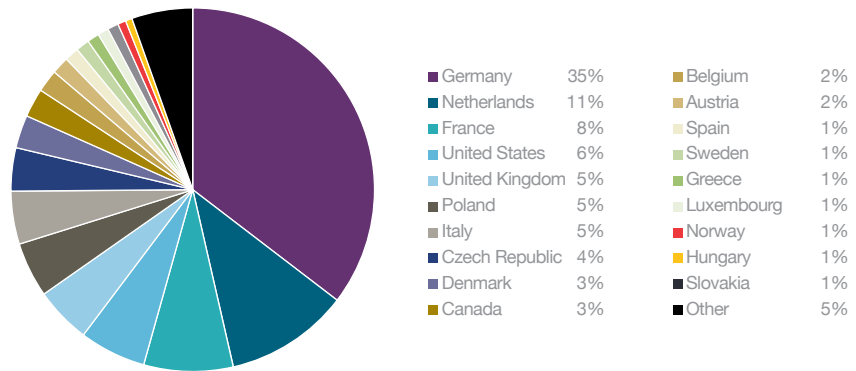
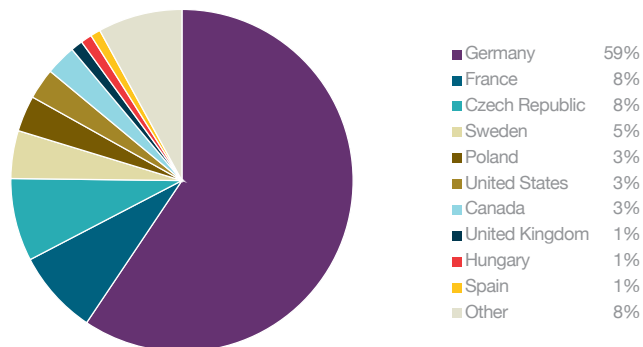


Figure 20 – .eu IDNs (Latin, Greek, Cyrillic script) by country of hosting




The results of the analysis of .eu IDN hosting support the findings of third party research which correlates local servers with local content, and emphasises the link which we have drawn in previous IDN studies between IDNs and content in the language indicated by the IDN script.

FOCUS B

Reading, visual perception and language competence: a complex relationship

Authors: Giovanna Marotta and Margherita Donati
University of Pisa



4 Reading, visual perception and language competence: a complex relationship

Authors: Giovanna Marotta and Margherita Donati
University of Pisa

4.1 The complex relationship between written language and visual perception


A growing number of studies reveal that spelling and reading are complex cognitive phenomena relying on sophisticated linguistic knowledge and involving both phonological and non-phonological knowledge. Proficiency in reading involves many variables: automaticity of word recognition, familiarity with text structure and topic, awareness of various reading strategies.

In this section, we describe the complex relationship between visual perception, writing systems and language competence. In particular, we focus on the interaction between native language (L1) and second language (L2) in reading. We explain why domain names should be in local languages and how multilingualism is necessary for an inclusive web.

4.2 How the brain perceives written language: (neuro-)cognitive studies

A current challenge within the field of cognitive science is to enhance understanding of how the brain processes letters and written language.

Behavioral studies into reading in native and second language have undergone a shift in perspective over the last forty years. Early theories considered the reading process to be a passive, bottom-up activity where readers reconstruct meaning from the smallest textual units. More recent models argue for an interactive perspective, involving both bottom-up and top-down processing, ie an interactive process between the reader and the text.



In the reading process, when visual information reaches the primary visual cortex, it is processed in order to establish whether or not letters are present. Letter perception is based on features, i.e. elementary and discrete visual distinctive elements. However, letters cannot be represented without reference to feature dislocation (i.e. gaps or space relations). Consequently, spatial relations of features must be taken into account.

Reading words is a cognitive activity that begins with processing the visual features of letters and ends with word recognition. When information reaches the word level, two different pathways (the lexical pathway, and the non-lexical pathway) contribute to perform reading. In the lexical pathway, lexical entries are recognised by matching them with the templates of our mental lexicon, which tries to match a written word against our own stored meaning for that word. At the same time, the non-lexical pathway is based on the grapheme-phoneme conversion mechanism that allows reading words that are not present in our mental lexicon (eg, new or fake words).

System and context elements show that perception is highly connected with human cognitive structures. Furthermore, understanding of both the context and knowledge of the meaning of individual words combine to play a relevant role in letter perception.

The notion of context is strictly related to the alphabet. The degree of knowledge of an alphabet (or various alphabets) is a crucial factor in the perception and categorisation of written language. For example, from a strictly visual point of view, the Latin letter < n > and the Greek letter < η > are confusable. To an individual who is familiar with both the Latin and Greek writing systems, these letters are less likely to be confused for one another than to an individual who only understands one of those writing systems. Obviously, the reader's degree of knowledge of the two systems will play a role in sorting out possible ambiguities. Relationships within a system are also relevant within the same alphabet. For example, the Latin small-case letters < n > and < h > are not ambiguous for a reader familiar with the Latin alphabet as the letters have quite different sounds and are rarely if ever used as a substitute for one another within words.

Another fundamental aspect is the interface with phonology. Phonological knowledge can impact letter perception and spelling performance. In particular, there are two core processes in the reading process: phonological decoding (ie the sound) and orthographic development (ie an understanding of systems of spelling). Phonological information not only plays an important role in early reading process, but continues to have a stable effect throughout it, challenging the view that more adult readers should rely less on phonological information than younger readers (eg “brane” as a prime for “brain”).



4.3 Processing a written second language (L2)


When learning a second language, the way people learn to spell is in part shaped by the characteristics of their native language and the writing system it uses. Literacy acquisition shows cross-linguistic differences between alphabetic and non-alphabetic languages. In alphabetic literacy, phonological competence (the ability to manipulate phonemes, syllables and their subconstituents, especially rhymes) plays a major role.

However, even within alphabetic orthographies, cognitive skills involved in literacy are not identical across different languages. Learning to process written language is easier (and faster) in transparent orthographies, ie where correspondences between letters and phonemes are consistent (eg Italian) than in opaque orthographies, where these correspondences are inconsistent, as in English or Danish. Native readers of inconsistent orthographies tend to read non-words using the large-unit strategy common in their native orthographic pattern, while for consistent orthography readers, only the phoneme level plays a role. For example, empirical evidence from English and German native readers shows that “English readers are forced to use at least two reading strategies: a small-unit strategy to reduce visual/orthographic complexity of large orthographic clusters and a large-unit strategy to reduce inconsistency that is maximal at the grapheme–phoneme level in English. In contrast, reading in German can be very successful using a small-unit grapheme–phoneme decoding strategy only, because of relative orthographic consistency at the small grain size level” (Goswami, 2013). These data suggest that specific alphabetic orthographies impact both the rate of literacy acquisition and cognitive strategies of reading and spelling in one’s native language.

Recent studies provide evidence that native languages’ orthographic properties regularly impact the process of any second language (L2) at written level. For example, different reading strategies relating to transparent vs. opaque orthographies affect reading and spelling in second language(s), since readers tend to transfer their reading skills from the native language into the second language.

As stated earlier, context strongly reduces ambiguity in word perception and recognition. Miller (2011), analysing native Chinese readers learning English as second language, points out that the second language reading proficiency may depend on the context (or lack of context) where words are read. The recognition process seems to be significantly impacted by incorporation of surrounding contextual information, resulting in findings that are quite different from those of studies of decontextualised words (Miller, 2011).

Moreover, semantic meaning plays a relevant role in word recognition: there is a difference between processing meaningful rather than meaningless or unknown words because patterns we do not understand are less easily processed. This is obviously relevant for readers of a foreign language.



Finally, in processing any second language familiarity plays a role too. Performance in recall is regularly demonstrated to be superior in the language with which individuals are most familiar. As a consequence, differences in the accuracy of recall can be identified in bilingual speaker/reader, depending on the level of competence of the language.

4.4 Concluding remarks

The findings from linguistics, cognitive psychology and neuroimaging provide evidence that any second language in its written form is processed at multiple levels and that the orthography of the native language impacts the processing of the second language.

Therefore, a truly multilingual domain name environment is the only way to make sure that each end user has the same rights to access the web in their native language from a technical and content perspective, and to experience it in full without the possible constraints and barriers of approaching it through a second language.



FOCUS C

Universal acceptance of IDNs

Author: Mark McFadden



5 Universal acceptance of IDNs

Author: Mark McFadden

Analysis of how the brain perceives language, especially in second languages, indicates a need for a multilingual domain name system. Particularly relevant are the relative difficulties in recalling words in second languages, and of the importance of context in enabling the brain to process written language. These insights suggest that internationalised domain names would offer Internet users great benefits in terms of memorability, understanding and intuitive signals of the type and language of content to expect.

So, how should we interpret the comparatively low uptake of IDNs to date? Does it signal lack of market demand or are other factors at play?

The next section reviews the issue of universal acceptance of IDNs – ie the ability to use them across all applications and services associated with domain names. It concludes that IDNs remain difficult to use and do not work at all in many contexts.

5.1 What does universal acceptance mean?

A fundamental feature of the “universal acceptance” of IDNs is how well the IDN can be used – in software, in mobile apps, in modern web-based applications and even in forms on paper or in PDFs. Traditional, ASCII domain names are used across many contexts, such as browsers, email clients, smartphone apps, online forms and accounts. For IDNs to fulfil their potential, they must not only be registered and available for use, but also able to be used wherever a more traditional ASCII domain name is used.

This principle is called *universal acceptance*. For the purposes of our report, universal acceptance is the relative level of ease of use, predictability and memorability of IDNs in Internet services and applications.



5.1.1 Universal acceptance, IDNs and recent experience

While universal acceptance is essential to the success of IDNs, it is also an issue with ASCII domains.

In the past, new gTLDs (such as .travel or .coop), failed to reach their intended audiences because of lack of universal acceptance. For example, popular software would reject domain names where the TLD was more than 3 characters.

If acceptance was a problem for the first rounds of new gTLDs (for instance, .info or .museum) then the problem is more complex in relation to IDNs. Many software developers, unaware that IDNs have existed for years, have built into their software functionality for checking on domain names, based on incorrect assumptions. These incorrect assumptions result in IDNs not responding to basic queries.

5.1.2 Universal acceptance and new internet tools

Too often, universal acceptance for IDNs is viewed only in the context of browsers. The word “universal” implies that IDNs should be accepted anywhere a traditional domain name might be. New development tools for the Web attempt to be “smart” about distinguishing between what is a link on a web page and what is simply text. For acceptance of IDNs to be truly universal, even our newest tools across a variety of platforms (traditional browsers, mobile devices, embedded devices in consumer electronics, etc.), need to be built with an understanding of how IDNs work.

Our examination of universal acceptance is, thus, broad. While we find that IDN support in browsers and other contexts is slowly improving – many challenges need to be overcome before universal acceptance of IDNs is achieved.

5.1.3 New domain names

Of course, the issues surrounding universal acceptance are not specific to IDNs. They will affect every one of the strings in ICANN's new gTLD programme. History shows that this is likely to be a large problem. While previous, small additions to the collection of gTLDs had few outreach efforts relating to universal acceptance, in the current round of new gTLDs, both ICANN and the underlying registry/registrar ecosystem recognise that universal acceptance is a critical issue – and both are committing resources to outreach and education.



5.2 Universal Acceptance and Mobile Devices

The success of both Wi-Fi and mobile broadband services has meant that there is a revolution in progress in Internet access and use. No longer is a traditional desktop or laptop computer the most common way to access the Internet. Instead, mobile devices – including tablets and smartphones – have become the predominant tool for accessing the Internet.

This trend shows no sign of ending soon. Recent projections show that the total number of high speed mobile broadband subscriptions is expected to grow to 1.3 billion by the end of 2018³⁴. In the same timeframe, there are expected to be 5.6 billion smartphones in service. There has been a 70% growth rate in mobile data traffic since our last report on IDNs³⁵.

Mobile devices and services have begun to dominate the way we think about access to the Internet. These changes are taking place all around the globe and universal acceptance for IDNs thus becomes a crucial issue for mobile devices.

5.2.1 Is mobile different?


The mobile Internet consists of two separate types of applications. Bespoke applications that are customised for the mobile environment and adaptive applications that are services originally written for traditional computers that have been adapted to work on multiple platforms including mobile devices.

Since adaptive applications are built upon traditional foundations, the applications share the same IDN usability challenges as the desktop/laptop applications do. We'll examine those traditional universal acceptance challenges for browsers and other Internet services in a later part of this report.

The industry for bespoke applications for smartphones, tablets and other mobile devices (for instance, wearables, and vehicle-based computers) is exploding. The three largest environments for bespoke applications are Google's Android, Apple's iOS and Microsoft's Windows Phone.

5.2.2 Built-in support

When a bespoke app developer needs to support IDNs on a mobile platform, they have two choices. They can build the code themselves, but this is error-prone and difficult to keep current with changing standards and developments in the IDN space. Their other option is to take advantage of standard code libraries which are common to all developers on each of the mobile platforms. In the past, this was impossible because the major mobile operating system vendors did not provide code libraries (called Application Programming Interfaces, or APIs for short).



Since our last report, each of the major mobile operating system vendors has produced IDN APIs that provide the essential tools for supporting the use of IDNs in bespoke mobile applications. This development means that if a developer so chooses, the tools are at hand to build Internet-ready bespoke applications that natively support IDNs. In interviews conducted for this study, both Google and Apple expressed an ongoing commitment to supporting those new APIs and keeping them up-to-date as standards in the IDN world changed.

In another development, an IDN API library for Android has emerged in the Open Source community. This is evidence that there is enough interest amongst developers to contribute the work to build their own, common set of tools.

5.2.3 Is the universal acceptance problem solved for mobile?

With the emergence of programming tools that support the development of custom, mobile applications that support IDNs, it would be tempting to say that the problem has been solved for mobile devices. Unfortunately, use of the APIs are optional, not automatic. A developer might choose to not use the APIs because it is easier to build the applications without IDN support. It's also likely that a developer might not even be aware of the existence of IDNs as they develop their Internet-enabled application.

While the emergence of APIs for mobile environments is a welcome development, the documentation of how to use the APIs is scant and examples of proper processing and display of IDNs were non-existent on all three major vendors' websites. Developers tend to use examples as an effective way to learn new skills. Without those examples, and the documentation to go with them, it will be hard for software developers to learn the skills needed to incorporate IDN support on mobile platforms.

Education is key to success in the mobile arena. An informal poll of a small number of app developers for Android indicated that the majority of them were staying away from supporting IDNs in their custom apps because they believed IDNs represented a security risk. The majority of those surveyed indicated that they worried that IDNs presented a string confusion risk to their users. Google has recently indicated that security and the potential for string confusion is the reason why IDN-enabled email addresses are not yet accepted as user identifiers.

As we see elsewhere in this report, security and string confusion can be addressed via policy rather than technology. Still, IDN implementations for mobile devices may be slow in appearing because developers are unaware of recent developments in policy-based protections for IDNs.



5.3 IDN Support in Web-based services

5.3.1 Overview

In last year's report, we emphasised the importance of IDNs outside the context of the browser address bar and inline hyperlinks. URLs are not just the foundation of the web; they play an essential role in the content and operation of the most popular services on the Internet. Domain names play such a ubiquitous part of the use of the Internet, that IDNs in browsers is a truly small part of the IDN Universal Acceptance issue.

If IDNs are to be used everywhere we can use a traditional domain name, then IDNs must be able to appear in user names, links to other resources, and in a variety of other places where the location of an Internet resource needs to be specified. In particular, popular social, blogging and photography sites make extensive use of domain names – and by extension, need to support IDNs.

There are two key factors for web-based services and IDNs.

First, the service should support IDNs just as they would any other URL that would appear in their services. In a social network, an IDN should be able to appear and be used, in the same way, as any URL created from a traditional domain name, eg in user generated comments, forums, or in reviews on Amazon or iTunes.

Second, if the service requires an email address as a component of the user identifier, the service should support email addresses built from IDNs as well as ASCII-based email addresses.

Our look at these two key criteria for success builds on our initial research from last year.

5.3.2 IDN URL support in web-based services - IDNs as content

One area in which there has been improvement in Universal Acceptance over last year is in the use of IDNs as content. If a service displays a URL, it should recognise that it is a link to an external resource and do the expected action when the text is clicked upon.

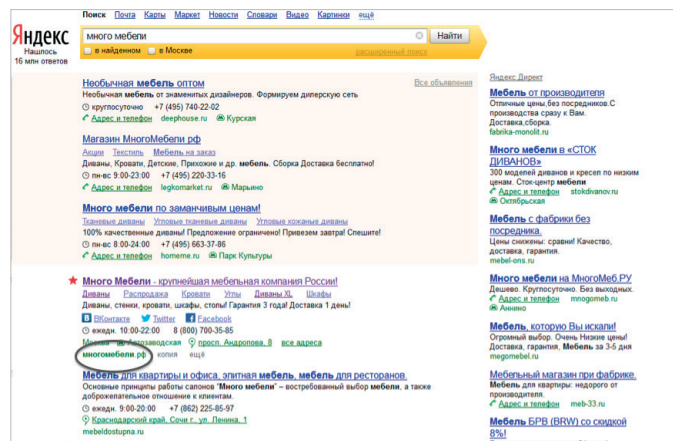
Domain names sometime appear in web pages as an integral part of the text. This happens often in social media and blogs. The domain name will simply be included as part of the text that the user is reading. In these cases, the browser must first determine that the text is a domain name and then decide how to handle the situation if a user clicks on the text. In the past year there has been a noticeable improvement in the way browsers handle domain names embedded in text.

Search engines and social media services have done the best job of transitioning to support IDNs. In the case of search engines, almost all support searching for, displaying and linking to web content that is located with IDNs. The expectation is that where a user's native language is supported for displaying the summary of the content, the link remains in

its native form (not converted to Punycode) and that the link works properly. We examined, where available, more advanced functions of search engines and found IDN support to be slightly more uneven than the underlying search tool. For instance, Google's ability to search on a particular top level domain works properly (for instance, searching for web-sites in the IDN TLD .онлайн) but other search engines failed similar tasks³⁶.

Search engines that are targeted at specific languages are beginning to show improvement in handling IDN search results. For instance, the Russian registry found that general purpose search engines were failing to promote Cyrillic URLs in their indexes. However, Yandex, the most popular search provider in Russian Federation, now offers IDN search results. For example, in the following screenshot, the circled domain name is многomeбели.рф.

Figure 21




Search engines are beginning to show IDN addresses in results

In the last year research in Republic of Korea found that browsers were able to successfully display the Korean 한국 domain names in all modern browsers. In addition, Google and local applications AL Tool and Naver are able to use Korean IDNs in the toolbar. Also in the last year, Naver and Daum (popular Korean applications) began to display 한국 (hanguk) domain names in search results³⁷.

Social media sites have also improved their support for IDNs as content. The two most popular by almost any measure, Facebook and Twitter, support and recognise IDNs as links. For instance, a Facebook posting which has an IDN included, automatically searches for and includes a snapshot of the referenced content – including the IDN in native form and a working link. The significant change over the last year is that many more social media services scan the text in a free form input box and are able to recognise and correctly process the IDN.

This improvement in Universal Acceptance is significant. Last year 92.3% of the sites we tested did not recognize IDNs in the same way as ASCII URLs. For this year's survey the number was 54.6% – a dramatic improvement in a single year³⁸.



This development shows awareness and improvement among some of the most commonly visited sites on the Internet. However, as we will see below, progress on Universal Acceptance is not universal.

5.3.3 IDN user identifiers

It remains conventional to use an email address as an identifier (or, user name) component of the security credentials that allow you to identify yourself to a website. While some sites give you a choice of an invented string, more than two-thirds (68.2%) of the sites we examined used an email address to identify a user of that service.

If Universal Acceptance is going to be a success, email addresses that are built out of a set of internationalised characters (for instance, андрей@проект.рф) can and should be supported as user identifiers in the same way as email addresses built from ASCII characters.

Last year we attempted to create accounts on eleven of the world's most popular services on the Internet using an email address that included an IDN. Many, if not all, of these services provide localised services in different languages across the globe. In all eleven cases, the account creation failed.

This year there has been no improvement. In fact, our survey in this area shows no change in all eleven of the top web sites visited by users who require a logon or security credential.

This appears to be a difficult problem for popular web sites: in cases where we found an improvement in the way that IDNs were supported in content, there were no corollary improvements in using IDNs in user identifiers.

None of the world's most popular sites allow users to create accounts with IDN email addresses

5.4 IDNs beyond the user

We have defined universal acceptance as the relative level of ease of use, predictability and memorability of IDNs in internet services and applications. In our annual survey we examine the acceptance of IDNs in browsers, email clients and on mobile devices. However, IDNs should – if the goal of universal acceptance is to be met – be able to be used in settings not so familiar to the average Internet user. In this year's report we examine two of those uses for domain names that are not typically encountered by a user, but still form an important part of how the Internet works: Digital Certificates, and Policy Data in the DNS.

5.4.1 Digital Certificates

Digital Certificates are an essential part of providing security on the Internet. Few Internet users encounter the certificates themselves, but nearly everyone relies on the services they provide. A digital certificate provides identity information about a person



Universal acceptance is just as important in services that the user doesn't see, eg digital certificates, policy data

or organization who presents it (called the subject), and the organization that vouches for the identity of the subject. In the physical world, some places require you to present a passport to prove your identity – with the government that issued the passport being the organization that vouches for your identity.

Digital certificates include the domain name of the “subject”. For instance, when making a secure (SSL or TLS³⁹) connection between the browser and website, the website presents the digital certificate to “prove” that they are who they say they are. The user's browser matches the domain name in the URL with the domain name in the digital certificate. If they match, the browser often shows a small padlock indicating a successful secure connection. If they don't match, or something else is wrong with the digital certificate, the browser shows a warning and may display a broken or open padlock.

What about the domain name in the digital certificate? Shouldn't it work the same way when an IDN is involved? The problem is: which domain? The native language IDN, or the one encoded into ASCII using Punycode?

The user expects that the native language one should be used. After all, this is what should be displayed in the Address bar of the browser. However, the browser uses the ASCII version of the IDN to find the web page to display. As a result, the two versions of the IDN domain name appear not to match, and thus the browser will issue its warning that the security connection is broken. The user may abandon the session based on that warning.


The alternative approach – of putting only the punycode version of the IDN into the digital certificate – is even less user friendly.

The solution that has emerged is for certificate authorities (the organizations that issue the certificates and vouch for the authenticity) to issue certificates that contain local language characters in all certificate fields. This means that the user sees precisely the same characters in the address bar of the browser as she does in the digital certificate. It is left to the browser – or, the underlying operating system – to make the conversion to the punycode format in order to execute the comparison of the domain name on the certificate and the domain name in the URL.

The ability for certificate authorities to support natural language is an important step in ensuring universal acceptance for secure communications over the Internet.

5.4.2 Policy data in the DNS

The Domain Name System (DNS) is a reliable database, distributed around the globe. It's natural that protocol designers look to the DNS when they want to make data available worldwide. The consequence of this is that the DNS does much, much more than simple conversions of domain names to IP addresses.



The success of the DNS means there can be quite a bit of information stored in each domain name entry. In fact, each bit of information stored in the DNS for a domain name is kept in something called a “resource record.” Having multiple resource records available for a domain name is a key part of the DNS’s flexibility.

Box 2 – DNS Policy data is essential for discovery of services

Policy data is frequently used to “discover” if services are available in certain networks. For example, a guest might bring a laptop into your house and connect it to your local area network for the first time. One of the things that the laptop is configured to do is to query the local DNS to see if certain services are available (for example, a printer, or support for its device drivers).

Having multiple resource records for IDNs is not a problem and is easily implemented. However, the storage of information in native languages, in the resource records themselves, is not possible. This means that IDNs can’t yet function in the same way as traditional ASCII domain names, in that IDNs can’t contain policy data which is “looked-up” by the DNS. This has implications for certain kinds of applications such as service discovery.⁴⁰

5.5 Universal acceptance and email

In the current year, the technical standards for Internationalised email have been finalised, bringing to a close a multi-year process.⁴¹ With the technical standards in place, it would be fair to consider the state of the Internationalised email market and the abilities of clients and servers to process Internationalised email.

A study by the Communications and Information Technology Commission in Saudi Arabia⁴² shows that email is the second most popular application (behind web browsing) when using IDNs. Those results show that extending support for IDNs to the internationalization of email is a natural and needed development.

Unfortunately, this year’s survey finds that universal acceptance of internationalised email is effectively non-existent. Just as in our report last year, no significant traditional, commercial email client or server has emerged to provide native support for the standards developed by the IETF. Just as it has taken browser developers some time to adapt to a world with IDNs, a similar delay is happening with email software developers. Given that this state of affairs has continued for years, it is worth examining the causes.



5.5.1 A difficult problem to solve

Internationalising email is a difficult problem to solve. The technical standards for email are so old that only simple ASCII characters were ever envisioned to be sent between senders and recipients. The simple act of including an “attachment” to email was, at one time, a profound problem to solve: how do you send a photo from a digital camera to someone who can only send and receive simple characters? The problem is similar for internationalised email addresses. How can I use characters from my native script if I only have access to very basic character sets?

The infrastructure for email is much more complex than most people realise. Internationalization of email addresses requires finding and supporting all the contexts in which email addresses occur. Email addresses appear in many places, including email headers and digital signatures. In every case, the servers – and the connections between them – must be transformed to accommodate internationalised addresses.

From the perspective of technical standards, the problem is profound enough that the entire landscape for email changed, including:

- a change to the fundamental standard for sending email,
- the ability to encode email message headers with UTF-8⁴³,
- new features to support internationalised delivery status and notification messages, and
- extensions to support internationalisation of both server based email (IMAP) and client-server email (POP)⁴⁴.

These major changes mean that internationalised email is, in a very important way, not compatible with existing email systems.

Incompatibility between Internationalised email and legacy systems is an enormous problem to solve. Even if your computer’s email client supports internationalised email, and you are connected to a service that supports internationalised email, it is still sometimes not possible to use that service. This is an indication of how difficult universal acceptance is going to be.

5.5.2 Implementation status for internationalised email

In the following table, we examine the status of Internationalisation for the top ten email clients across all platforms⁴⁵. In previous years we separated mobile clients from desktop clients and analyzed each separate from web-based clients. In addition to this analysis (see below), we also recognise that these clients are beginning to merge across platforms. We decided to use statistics that find the most popular clients and then examine the state of internationalisation of each.



Table 5 – Support for internationalised email addresses in top ten email clients across all platforms

Market position	Client name	Share of market	Supports	
			International Email Addresses? ⁴⁶	Sending of International Email? ⁴⁷
1	Apple iPhone	26%	No	Yes
2	Outlook	14%	No	No
3	Apple iPad	12%	No	Yes
4	Gmail	12%	Yes	Yes
5	Apple Mail	8%	No	Yes
6	Google Android	6%	No	Yes
7	Outlook.com	6%	No	No
8	Yahoo! Mail	5%	No	No
9	Windows Live Mail	2%	No	No
10	Windows Mail	2%	No	No

Gmail now supports IDN email addresses

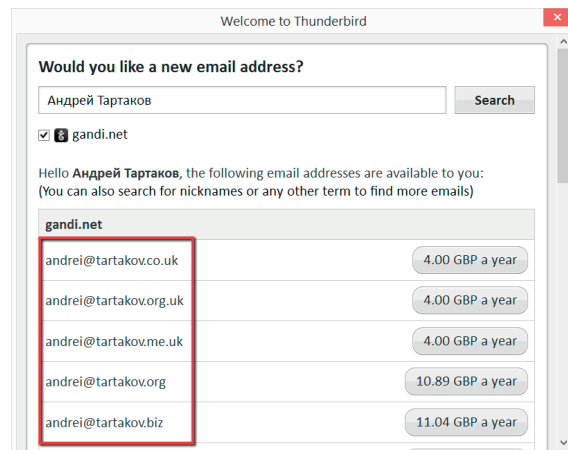
5.5.3 Email clients – traditional clients

Traditional clients such as Microsoft Outlook, Apple Mail and Thunderbird from Mozilla all succeed at displaying email in the local language in which the email was composed.

While welcome, this is different from Internationalised email. With internationalised email there needs to be support for composing mail headers with IDNs and Internationalised email addresses as well as being able to send the message to the nearest server. It is in this area that there is near total failure to Internationalise email.

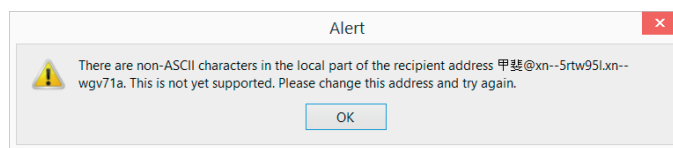
As an example, when using the email client Thunderbird, users are given a chance to set up a new email account. However, it can't be done with an address that supports internationalisation:

Figure 22



Since our last survey of traditional email clients, there has been little progress in deploying support for Internationalised email. For instance, attempting to use Thunderbird to send a message to an internationalised address results in this error:

Figure 23

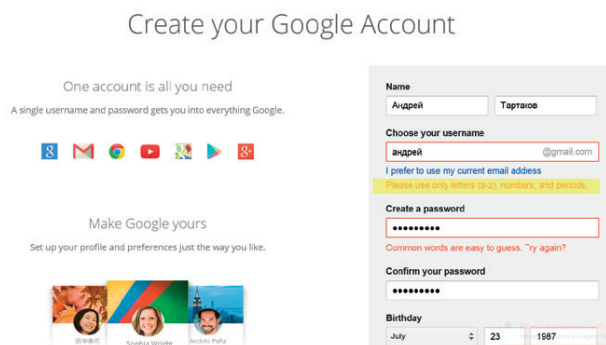


However, we have seen experimental support for internationalisation through third-party plug-ins to traditional software (for instance, Raseel for Arabic support in Outlook).⁴⁸ It is also worth noting that COREMAIL, the commercial email system with an extremely large market share in China, already supports internationalised email⁴⁹. CNNIC has also announced their intention to help drive the development of the internationalisation of Postfix, an open source email server, by the end of this year⁵⁰.

5.5.4 Email clients – web-based services

Web-based mail (such as Gmail or Yahoo! Mail) continues to be very popular. It is easily localised so that the body of email messages can appear in local languages. However, with almost no exceptions, it is impossible to use Web-based services to set up a new email account using an Internationalised email address. Here is a screenshot showing what happens if you try to set up a new Gmail account with an internationalised user identifier:

Figure 24



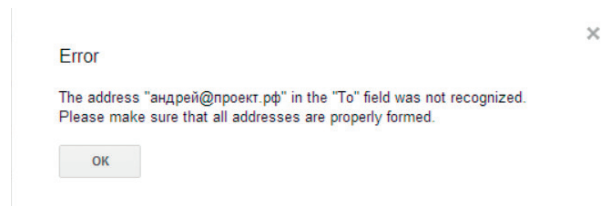
In addition, while it is easy to compose the body part of an email message in a local language and have it delivered properly, it is almost uniformly impossible to address that same message to a recipient that has either a fully internationalised email address or one where the email address is ASCII to the left of the “@” sign and an IDN in the domain name part.



Here is an example from Gmail, indicating that it is unable to accept an internationalised email address:

Gmail users are amongst the first to be able to use a global, web-based email service that supports internationalisation

Figure 25



Since our last survey of web-based services, we have identified only 2 of 15 services that have any support for Internationalisation. Recently, Google has successfully deployed a new version of Gmail that supports Internationalised Email. This means the Gmail users are amongst the first to be able to use a global, web-based email service that supports internationalisation.

5.5.5 Email clients – mobile and portable devices

In the past year we have seen a major change in the operating systems for portable and mobile devices. For all three of the most popular mobile operating systems, programming libraries have emerged that allow programmers to support the manipulation of strings that make possible support for IDNs.

Regrettably, this does not mean that programmers have the tools they need to support the development of email clients that are fully internationalised. In particular, programmers do not have the tools they need to insert non-ASCII strings properly in email headers.

In our tests, with the Apple iPhone, an Android based phone, the Amazon Kindle and a Windows 8 Phone we found no support for an email implementation that allowed for Internationalised accounts or the standards-based sending of Internationalised email.

5.6 IDNs and browsers

5.6.1 How a browser decides which IDNs to display

Browsers face a choice of what to display in the address bar when displaying an IDN. With well-known homograph attacks⁵¹, browsers are particularly sensitive to displaying the native version of the IDN. Because of this, almost all browsers display the punycode version of the IDN in the address bar rather than the native IDN. This is a problem for universal acceptance, so most of the major browser developers make display of the IDN conditional based on a series of criteria we examine in the following table.



At the time of the research for this study, we found the following:

Table 6 – Status of IDN support in popular browsers

	Google Chrome	Microsoft Internet Explorer 11	Mozilla Firefox	Opera	Safari
Can be forced to always show the IDN URL?	No	No	No	No	No
Decides whether to show the IDN URL as a whole or label by label? ⁵²	Label by label	Label by label	Label by label	Based on the TLD only	Based on the script only
Contains a blacklist of characters that will prevent display of the IDN URL?	Yes	No	Yes	No	No
Has a configurable list that will allow display in specific languages?	Yes	Yes	No	No	No
Has a whitelist of TLDs and will only show the IDN for TLDs in the list?	No	No	Yes	Yes	No
Has a whitelist of scripts and will only show the IDN for scripts in the list?	No	No	Yes, with algorithmic exceptions	No	Yes
Allows for hybrid IDNs such as <code>http://www.research.онлайн?</code>	Yes	Yes	Yes	Yes	Yes

Browser support for IDNs is patchy, but improving

The approach taken to the homograph attack problem is not the same for each browser. The result is two-fold: first, the IDN experience changes from browser to browser with the same IDN and website; and second, several browsers require user configuration before IDNs can be displayed. For people who use two browsers there is an inconsistent experience that creates confusion and reduces trust.

There are policy implications related to individual browser developers using a white list of TLDs as the primary basis for deciding whether or not to display the IDN. This means that IDNs whose domain registries have put in strict controls on applied-for domain names are able to have their IDNs displayed automatically. The browser developer is the source of judgment of whether the registry policy is appropriate and there is no mechanism for appeal.

Modern browsers for the desktop and laptop have fully compliant support for IDNs, but while that is necessary for the success of IDNs on the World Wide Web, it is not sufficient. When a user sees the punycode (as in so many situations in the table above), the URL displayed is devoid of meaning. As a result, the user is less likely to use it as a bookmark, send it to a friend in an email message or post it as a link. In other words, the default display of the URL in Punycode strips the URL of any of its semantic value – and, as a result, makes it significantly less usable.



5.6.2 IDNs and the mobile browser

For this study we repeated our examination of IDN support in browsers running on portable, mobile and embedded devices. In the last year, we have seen upgrades in the sophistication and capabilities of each of these browsers. On iOS, versions of Safari, Chrome and Opera have each been upgraded. Internet Explorer remains the natural choice for Windows Phone and Chrome is the choice for browsing on Android devices. The question is: can IDNs be supported in these smaller, portable devices in the way that they are supported in traditional laptops and desktops?

In the research conducted for this report we found that Internet Explorer on Windows Phone, Chrome on iOS and Android, and Safari on iOS could all successfully display IDNs. In addition, the mobile versions of the browsers reflected the IDN display decision-making algorithms outlined above. Each manufacturer has made a conscious decision to avoid homograph attacks no matter what size platform the browser runs on.

Last year we noted that the user's ability to type an IDN into a mobile browser was severely limited. Especially for smartphones and embedded devices, the limitations of screen space and character count was obvious. Over the last year, both device manufacturers and third-party software developers are making significant improvements in this area, for example in virtual keyboard panels.

5.7 Conclusion

Significant progress has been made since last year in the emergence of standardised programming tools to support IDNs, and the recent announcement that Google is supporting IDN email addresses in Gmail. However, much remains to be done before universal acceptance of IDNs is achieved. The past twelve months has seen no meaningful progress in enhancing universal acceptance in browsers, email clients and software, or web-based social services.

Universal acceptance means more than just browsers and apps – in the last year, more examples of places where IDNs do not function in the same way as traditional ASCII domain names have emerged.



FOCUS D

Industry opinions

6 Industry opinions

6.1 A reminder of the domain name supply chain

The different actors within the domain name supply chain all have similar names.

The **registry** is the operator of the Top Level Domain, and is responsible for maintaining the database of all domain name registrations and their associated IP addresses (equivalent to a wholesaler). The registry is the authoritative entity for that Top Level domain, and is included in the “root” directory, the Internet Assigned Numbers Authority (IANA) database. Most of the larger registries included in this study do not have direct interaction with end users at the domain name registration phase.

Registrars sell domain name registrations to end users (retailer). They typically also provide a number of other services, and many offer a range of different TLDs to their customers. Registrars are usually accredited or otherwise authorised by a registry to sell individual TLDs.

Registrants are the people or organisations who register domain names for their own use (customer).

Figure 26 – The domain name supply chain





6.2 Methodology – registry survey

Since 2011 we have sent out an annual questionnaire to registries. In previous years, the survey has been sent to CENTR members and associate members, and to individual registries with whom we are in contact. This year, thanks to the collaboration of the regional ccTLD organisations APTLD and LACTLD, the survey was sent to ccTLD registries in the Asia Pacific and Latin American regions.

The survey asks for opinions on four questions:

- 1 How does uptake of IDN registrations relate to your expectations?
- 2 How well are IDNs supported by your registrars?
- 3 How would you rate end-user awareness of IDNs
- 4 What single change would improve IDN uptake?

Each question was scored on a Likert scale⁵³ from 0 (far below expectations) to 5 (exceeds expectations).

The first question aims at identifying any gaps between the level of IDN registrations, and the registry operators' expectations. A low uptake may be completely in line with expectations, for example, when there is a low population of people using the relevant character set in the target market.

The second and third questions are aimed at the two primary methods of sales: in marketing jargon, one is supplier “push” and the other is end user “pull”. If registrars (the channel to market) are not able to support IDNs, then a marketing **push** (eg through advertising, price promotions or other push strategies) will not be effective. Likewise, if customers are not aware of IDNs, then there will be little or no consumer **pull** (eg proactive requests by customers).

The fourth question is aimed at identifying the perceived barriers to greater uptake of IDNs.

This year, 58 registries responded to the qualitative questions, the largest response so far. The registries represented a geographically diverse sample including Europe and North America, Latin America, Arab States, and Asia and Pacific. Not every registry answered every question.

The low numbers in the data set can lead to potential distortions in percentage differences. However, the participants are expert in the field, and manage a large portion of the world's domain names. So, while the results are not conclusive, they give an interesting picture of industry impressions of IDN uptake, from a geographically diverse base.

A note on the data:

Last year, we stated the results of this survey as 2010-2013 (ie the years in which the surveys took place). This year, we have restated the dates to reflect the fact that these questions are part of a larger survey which asks for information about the year just past. The numbers of responses in past years are slightly increased compared with those published in previous years, and therefore small anomalies in the stated percentages may occur. However, the general trends are unchanged.

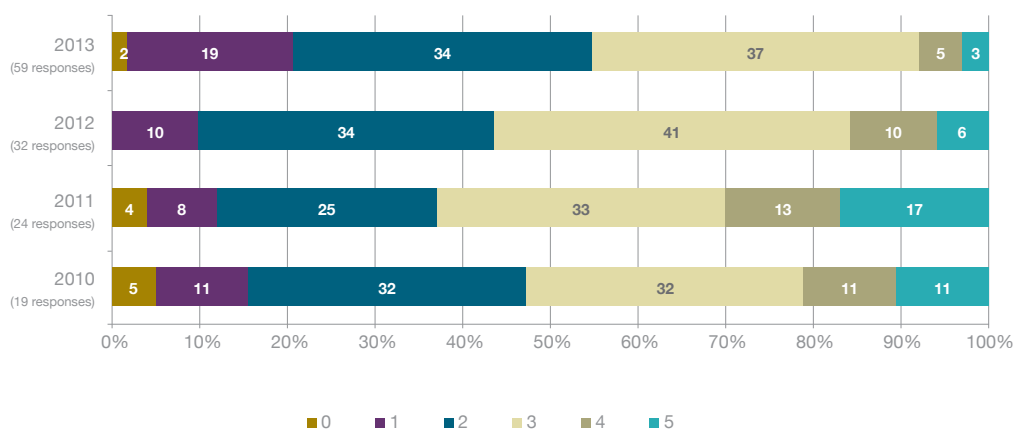
6.3 Results of registry survey

6.3.1 How does the uptake of IDN registrations relate to your expectations?

Only 8% of registries think that uptake of IDNs is very good or exceeds expectations

Over time, there has been a decline in registries' opinions about the level of uptake of IDNs in relation to expectations (figure 27). The average for 2013 has declined to 2.3 from a high point of 2.9 (2011). This suggests that registries are comparatively less happy with the level of IDN uptake than in previous years. Only 8% of registries who responded to the survey told us that uptake of IDNs was very good or exceeded their expectations, reduced from a high point of 30% in 2011.

Figure 27 – How does the uptake of IDN registrations relate to your expectations? (0-5) 0 = below expectations, 5 = exceeds expectations

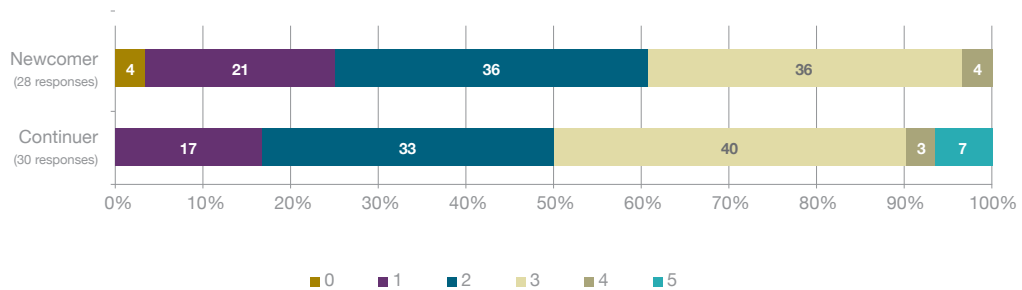


At the other end of the scale, 21% of registries indicated that uptake of IDNs was below or far below expectations. This is the strongest negative score in the four year survey, suggesting that confidence may be declining.

In all, the negative scores are increasing and the positive scores are declining.

For 2013 we had a much larger response rate than previously, thanks to the cooperation of LACTLD and APTLD. We have broken out the results for 2013 into “newcomers” and “continuers” to see whether the new responders had an effect on the overall data (figure 28). To some extent, they did. The two most negative scores were 8% higher for newcomers than continuers, and the two most positive scores were 6% lower. However, even excluding newcomers, the average scores for 2013 are 2.5 (a drop of nearly 0.5 since last year’s survey).

Figure 28 – How does uptake of IDN registrations relate to your expectations? Results for 2013 by newcomer and continuer



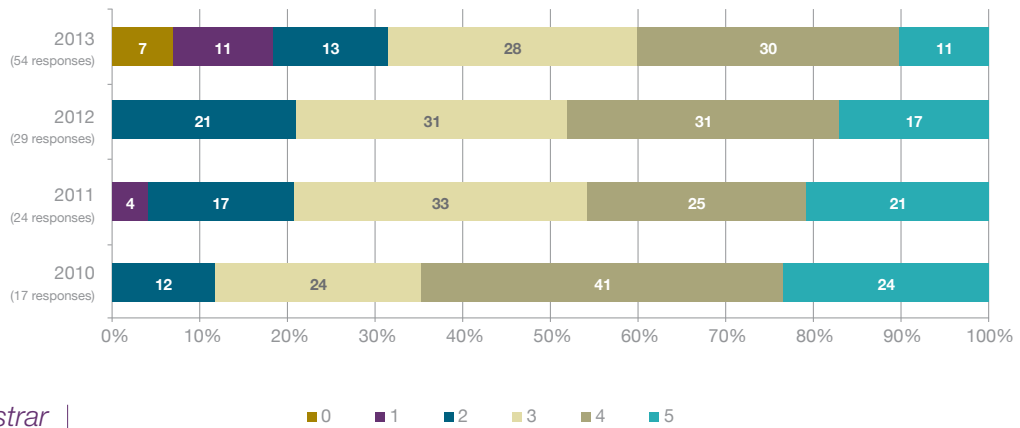
6.3.2 How well are IDNs supported by your registrars?

In previous years, registries have generally been upbeat about registrar support for IDNs, with average scores comfortably above 3.4 for the past 3 years. For 2013, the average score has dropped to 2.9, indicating that registries are less satisfied with levels of support for IDNs by registrars than in previous years (figure 29).

41% of registries who responded to the survey told us that registrar support for IDNs was very good or exceeded their expectations. This represents a 7% decrease since last year’s survey, when 48% of registries who responded to the survey rated registrar support as very good or exceeding expectations (46% in 2011, and 65% in 2010).



**Figure 29 – How well are IDNs supported by your registrars?
(0-5) 0 = below expectations, 5 = exceeds expectations**

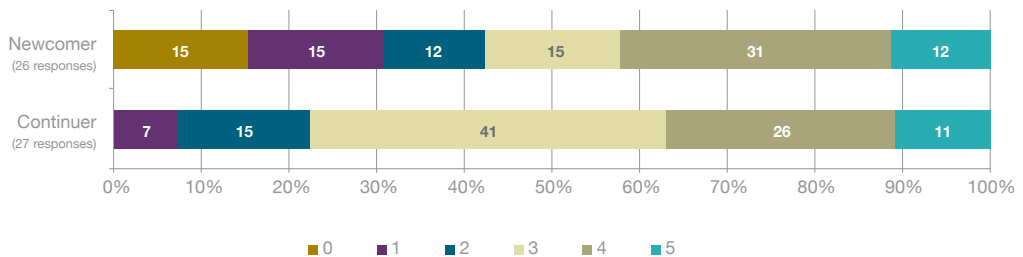


41% rate registrar support for IDNs as very good or excellent

This year, 18% of registries who responded used the two lowest ratings, compared with 0% last year. This is the largest percentage of lower scores since the survey began, and is 14% above the historical lowest score (4% in 2011).

Again, the number of responses received in 2013 was much higher than previously, so we reviewed the answers by newcomers and continuers to see whether the new responders affected the outcome (figure 30).

**Figure 30 – How well are IDNs supported by your registrars?
Results for 2013 by newcomer and continuer**



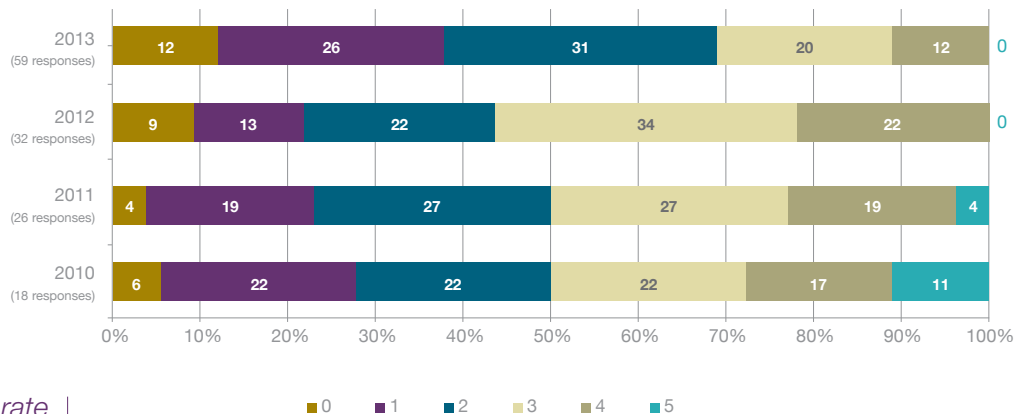
The 2013 newcomers, mostly from the Latin America and Caribbean and Asia & Pacific regions were on average 0.6 more pessimistic than continuers (2.6 versus 3.2). As a whole, the main difference is in the two middle bands, where the newcomers score 29% lower than the continuers. The two most negative scores were 23% higher for newcomers compared with continuers (none of the continuers used the lowest score), but the highest two scores were also higher by 6%.

After discounting the effect of newcomers, the average score (3.2 in 2013) is only slightly lower than in previous years.

6.3.3 How would you rate end-user awareness of IDNs?

The gloom continues in the registries' rating of end-user awareness (figure 31), traditionally the poorest performer in the group of three questions. For the first time, the average has dropped below 2, (1.9 in 2013, compared with around 2.5 in previous years).

**Figure 31 – How would you rate end-user awareness of IDNs?
(0-5) 0 = non existent, 5 = excellent**



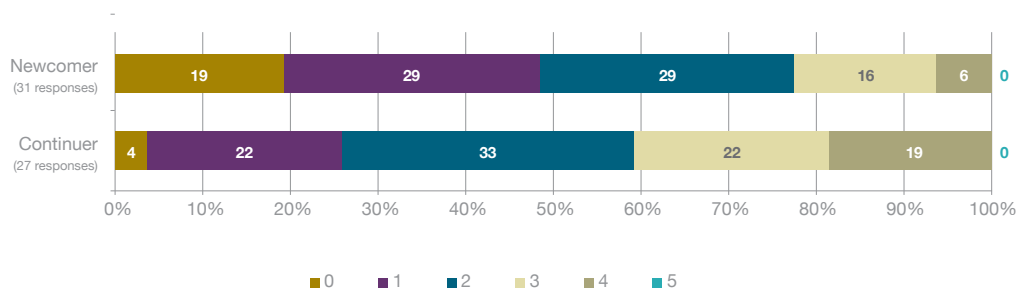
38% of registries rate end-user awareness of IDNs as below or far below expectations

12% of registries who responded to the survey told us that user awareness of IDNs was very good and 0% felt that it exceeded their expectations. This is a 10% decline for the top two categories compared with 2012, and significantly lower than the highpoint of 24% in 2010.

At the other end of the scale, 38% of registries indicated that end user awareness was below or far below expectations (compared with 22% in 2012, 23% in 2011, and 28% in 2010), with an increase in the “below expectations” category to 26% (compared with 13% in 2012).

Again, newcomers had the effect of depressing the scores compared with continuers (figure 32). The average score for the continuers is only marginally less than it was last year (2.3 for 2013 continuers, compared with 2.5 last year).

**Figure 32 – How would you rate end-user awareness of IDNs?
2013 newcomers compared with continuers**



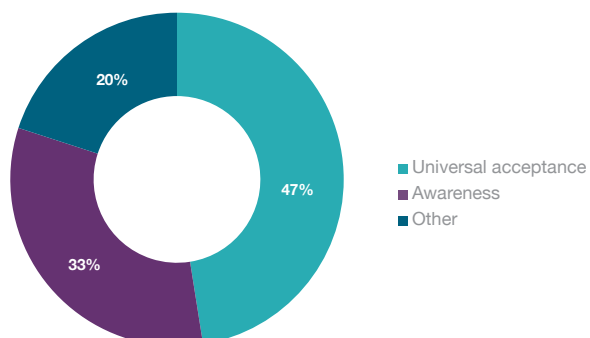
6.4 What single change would increase uptake of IDNs?


Every year, our registry survey participants are asked what single change would improve uptake of IDNs. The responses are free-text, and no suggestions are given.

This year, we had 40 responses to this question, more than double the response rate in previous years. Despite these, the responses highlight similar topics in similar proportions across the years (figure 33).

**Figure 33 – What single change would you make to increase uptake of IDNs?
2013, 40 Responses**

Experts believe that universal acceptance and improved user awareness would improve IDN uptake





As in previous years, two broad themes come through from the responses: the need to improve universal acceptance, ie support for IDNs across browsers, email, and internet applications (47%) and improvements to user awareness through marketing, or price promotions (33%). Other comments (20%) included the need to liberalise the registry's registration policy, adoption of the most up to date Internationalised Domain Names' technical standard (IDNA2008), and an increase in local content.

6.5 Registrar survey

This year, we also repeated our survey of EURid registrars. The registrars were selected by EURid from its accredited registrar base, with regional balance, and different business models. The survey was completed by 18 registrars, compared with 23 last year. All of the registrars surveyed offer IDNs to their customers. All of them offer IDNs under .eu, and many also offer IDN registrations under other ccTLD and gTLD extensions.

Registrars operate closer to the end user in the supply chain compared with registries, and several in our survey sell across many TLDs. Therefore registrars' opinions are likely to be informed by their knowledge of how IDNs are performing compared with the registrars' other domain name offerings.

With such a limited data sample, small changes in numbers can produce large percentage differences. Comparisons are difficult to make with confidence, as not only are numbers of responses lower, but also the surveyed registrars are not the same as last year.

Nevertheless, the trend is clearly downward. As with last year, registrar opinions were more pessimistic than their registry counterparts about both uptake and end user awareness of IDNs.

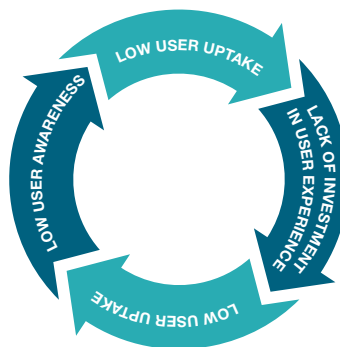
- In response to the question “how does uptake of IDN registrations relate to your expectations?” the average score has decreased to 2.0 (a decline of 0.6 compared with last year's 2.6). There were no responses at all in the two highest categories (“very good” and “exceeds expectations”)
- The average response to the question “how would you rate end user awareness of IDNs” dropped to 1.9 (0.4 less than 2012). Again, there are no responses at all in the two highest scores.

The survey also indicated that registrars continue to offer the expected range of services for IDNs (email forwarding, webhosting), and rely primarily on “push” marketing strategies for IDNs, ie reliant on customers to initiate the enquiry rather than, for example, advertising IDNs on their home page or elsewhere.

6.6 Conclusions

Our 2013 registry survey had more responses than ever before, thanks to the cooperation of regional ccTLD organisations. This has improved the geographical balance of experts in ccTLD domain name management, who manage many millions of registrations. The response rate to the registrar survey was lower this year because the registrar market has undergone major changes in the past months and therefore, it was more difficult to collect sound responses.

Figure 34 – IDN negative cycle



Nevertheless, there does seem to be a downward trend in levels of confidence relating to IDNs, whether it is uptake of registrations, support by registrars or end user awareness.

We suggested last year that IDNs may be suffering from a negative cycle (see figure 34). Without universal acceptance across applications and email, the user experience of IDNs remains poor. Users are lacking incentives to use IDNs, and therefore registrations are not yet achieving their potential. The knock-on effect is that many Internet users are simply unaware of IDNs, even those in countries and territories where one would expect a high uptake.

Confidence in a product or service within the supply chain is essential for success, and it is hoped that advances in universal acceptance will restore confidence to previous levels.

The results of both the registry and registrar surveys highlight the need for regular and even stronger cooperation between these two players in the domain name chain (registries and registrars) to ensure greater adoption of IDNs at the end user level, thereby supporting linguistic diversity online.

FOCUS E

IDNs in Arab States



7 IDNs in Arab States

This report has demonstrated that there is a clear link between IDN script and the language of web content, and highlighted the clear benefits to individuals of being able to read and understand domain names in their native languages. It has also highlighted that there are significant challenges to using IDNs (universal acceptance), which are evident in the responses to our industry questionnaire. Although progress is steadily being made, these challenges go to the core of IDNs' functionality, have inhibited uptake, and will likely continue to do so in future unless addressed.

Therefore, it seems that IDNs have the potential to foster and signal the presence of multilingual online content, but IDNs are currently underachieving compared with their potential. We believe that, given the strong linkages which we have shown between IDNs and language of content, that underachievement poses a risk to the successful migration towards a truly multilingual Internet.

It is clear that IDNs in some countries or territories are doing better than in others. In an effort to understand why this might be, we started to gather information about local conditions through country case studies. Over the past three years, we have developed case studies for nine countries that have implemented IDNs at the top level: Russian Federation, United Arab Emirates and Islamic Republic of Iran, Qatar, Saudi Arabia, Egypt, Republic of Korea, China, Viet Nam.

The country case studies were made possible through the generous collaboration of ccTLD registry staff in the countries.

The countries were selected because, unlike many of the European ccTLDs which operate primarily in Latin script and use IDNs to represent special characters (eg é, ç, ö, å, ø), the countries of the case studies are not well served by the mixed-script, hybrid IDNs. The exception is Viet Nam, which uses Latin script, but has been included because of its extraordinary experiences in implementing IDNs.

The ccTLD registries in the countries have played an active role in advocating the adoption of IDNs at the top level, and have been first movers in rolling out fully internationalised domain names.

7.1 IDN Readiness matrix


Each of the countries in the case studies was evaluated across numerous indicators, which have remained consistent with last year and build upon research by ISOC, UN-ESCO and OECD⁵⁴. These seek to identify the “IDN Readiness” of a country or territory, as follows:

- Country/Language factors
 - Level of linguistic and cultural homogeneity
 - Presence of Local Internet Exchange Points (IXP)
 - Broadband penetration (fixed and mobile)
 - Local language content
 - Size of population, and online population
- ccTLD (local domain name) factors
 - Strength of local registrar network
 - Registration policies
 - Low prices
 - Strength of ccTLD brand

The IDN Readiness matrix was presented in the World Report on IDN Deployment 2012, which set out the rationale for each of the indicators. These include the significant correlation that has been identified⁵⁵ between the development of network infrastructure and the growth of local content. The measure of “cultural homogeneity” is built up through a range of indicators on cultural diversity⁵⁶, net migration both in the general population and of students, international flows of selected cultural goods and services and where country data is present highlights from the World Values Survey⁵⁷.

Figure 35 – IDN-readiness matrix ALL





The results are presented in the IDN readiness matrix (figure 35), and help to explain why IDNs are doing comparatively well in some countries and not in others. It must be emphasised that this analysis makes no judgment on the countries or territories, ccTLD registries or any aspect of their operation.

The vertical axis reflects the summary of country/language factors. These reflect the macro-environment in which the IDN is offered. Some of the factors (linguistic homogeneity, cultural homogeneity) will be slow to move; others (broadband penetration, presence of IXPs, online population) can change quite rapidly, enabling mobility through the vertical axis.

The horizontal axis reflects the summary of the micro-environment that is the ccTLD registry, its policy, pricing, brand and crucially a network of local registrars. These are more readily affected than the country factors, and therefore it is foreseeable that individual countries or registries could have high mobility across the horizontal axis year by year. There are no significant movements from last year's chart, as there have been no major policy changes in the registries within the sample.

Appendix 2 sets out the tables with the IDN Readiness indicators, and a description of the IDN experiences across the nine countries studied.

7.2 Focus on Arab States

Last year's report focused on the experiences of the Asia and Pacific IDN implementations. Cultural, linguistic, Internet and ccTLD indicators are all favourable to IDN deployment. However, even in the Asia and Pacific region, overall the IDN registration numbers remain below the level of ASCII registrations in most cases, and are far from achieving their potential.

This section returns to the Arab States which were last reviewed in detail in the 2012 edition. Since then, Egypt has opened .مصر for general registrations. This section gives an overall view of Internet and social media uptake in the region. It then reviews implementation of IDNs by country or territory, with a particular focus on Egypt, Saudi Arabia, Qatar, United Arab Emirates, followed by a round-up for the rest of the region. It highlights challenges specific to the region. Next, we review how Arabic IDNs are being used, and find the familiar strong correlation between the language of web content and the script of domain names. Where local language content is present, it is typically associated with "high involvement" content (such as online business sites or blogs) – in contrast to "low involvement" content (such as pay-per-click, parking and starter sites) where the language is not one that uses the Arabic script of the domain name. Local language content associated with Arabic script domain names is also more likely to be hosted locally. The section reviews other developments in the region, including the liberalisation of registries in United Arab Emirates and Qatar. It finishes by noting the important work of the Task Force on Arabic Script IDNs, which is working to achieve standardised rules for generating Arabic domain name labels, handling of variants and registration data.



7.3 Overview – growth of Internet in Arab States

There are 152 million internet users across Arab States, a 41% internet penetration rate across the region⁵⁸. Use of social media is growing rapidly in the region, with 55 million Facebook users in May 2013 (compared to 45 million in June 2012). Egypt has the highest number of Facebook users in the region (13 million). Meanwhile, Facebook use has decreased sharply in Saudi Arabia, Bahrain, Kuwait, Lebanon, Qatar and UAE.

There are nearly 4 million Twitter users across the region, an increase of 79% from 2012 to 2013. Saudi Arabia has the highest number of Twitter users (1.9 million), an increase of 128% on the previous year.

Arabic is growing as a preferred language used by Facebook users, particularly in Yemen (81%), Palestine (75%), Libya (71%), Iraq (67%) and Egypt (61%). Use of Arabic as the language of tweets from the region is also on the increase, with 90% of Saudi tweets (and increase of nearly 20% on 2012), 83% Kuwait (increase of 5% on 2012), and 65% of Egyptian tweets (increase of 16% on 2012).

The Dubai School of Governance has analysed the language of government tweets in the region⁵⁹, and finds that, with the exception of UAE and Qatar over 90% of tweets from governments are in Arabic. Users tweeting about their government ministries also tend to use Arabic.

There are 4.7 million LinkedIn users across the region (compared with 4.3 million in 2012). In the region, 285 million videos are viewed every day from YouTube, and more than two hours of video are uploaded every minute. Saudi Arabia leads the region in YouTube playbacks, followed by Egypt, Morocco and UAE⁶⁰.

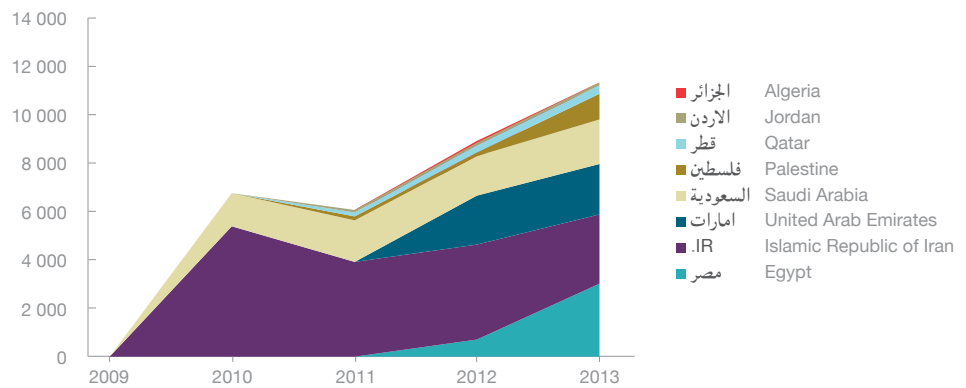
7.4 IDN adoption in Arab States

Across the Arab States, eight countries have deployed IDNs (Oman having deployed during 2013-2014). New gTLD IDNs have also come on to the market, including dotShabaka (شبكة) and .bazaar (بازار). Even though Arab States have been active in advocating the implementation of IDNs at the top level, and at the cutting edge of early implementation of IDNs, experiences in the region have been mixed.

According to our figures, there are now 12 000 Arabic script IDNs in ccTLDs from the Arab States. This includes both fully internationalised (IDN.IDN) and hybrid (Arabic.ASCII) domain names (figure 36). While the numbers are low, the percentage growth rates are high, even allowing for shrinkage of some registries (the Iranian registry has not accepted new registrations for IDNs under .ir since 2010, and the number of registrations has halved since then).

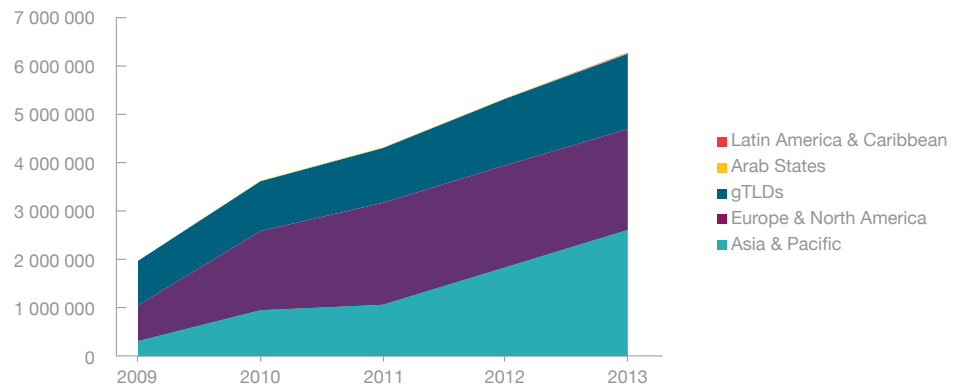


Figure 36 – Arab States – IDNs 2009-2014



For some years, it has been apparent that IDN registration volumes are not yet fulfilling their potential across Arab States

Figure 37 – IDNs by geographical region (both hybrid and full IDN) 2009-2013



For some years, it has been apparent that domain name registration volumes across the Arab States are not yet fulfilling their potential. For example, a review of IDNs by geographic regions (and gTLDs) shows that Arab States⁶¹ have only 0.2% of the world market.

The low uptake in domain name registrations in across the Arab States can partly be explained through our IDN Readiness Matrix (figure 35). In some countries, such as United Arab Emirates and Qatar – despite progressive registry policies and excellent Internet infrastructure – the populations are low and extremely diverse (both countries



Specific challenges for Arab States include immature domain name market, and related industries

have high levels of immigrant workers, and English is commonly spoken). Elsewhere, there are higher populations, greater linguistic homogeneity but other factors (for example strict registry policies, low instances of local language applications, no Internet Exchange Points) tend to inhibit mass uptake.

The case of Egypt is puzzling, especially as online populations are growing in numbers and vigour, and as there is thriving local language content. In Egypt, low volumes of domain name registrations (both under .eg where there are 9 000 registrations, and *مصر*.) contrast with mass adoption of Facebook (12 million Egyptian users), a large online population (44% of Egyptians are online), and the popularity of Arabic language content⁶².

The potential for mass uptake in Egypt is high, and this is reflected in our IDN Readiness Matrix (figure 35). Indeed, for the region, Egypt is already market leader for IDN uptake despite only opening for landrush during 2013.


7.4.1 Specific challenges in the region

The Egyptian registry highlights⁶³ specific challenges facing growth of IDNs in the region:

- Immature domain name market, and related industries
- Lack of awareness and outreach
- Complex registry policies
- Lack of universal acceptance means a poor user experience for the Arabic speaker
- Lack of appropriate skills and experience in the region (capacity)

Some countries with higher IDN adoption (eg Russian Federation, China) had been starting from a baseline of large numbers of ASCII registrations, and local registrar networks. In Egypt – and across the Arab States – there are low numbers and low growth rates of ASCII domain names. Recent years have seen the re-launch of a number of ccTLD in the region following reorganisations. This means that the populations may not be accustomed to seeing local domain names (we have called this “registry brand” in our IDN Readiness analysis).

Registries across the region highlight the lack of local registrars, and many have been working to recruit registrars (eg Qatar, UAE, Oman). At the moment, with the exception of the UAE, current rates of demand are insufficient to sustain a local, profitable industry until critical mass is reached. Registries report that local populations tend to use Facebook pages and search as substitutes for domain names. The Egyptian registry suggests that new market offerings are needed, which provide turnkey solutions for different market segments (doctors, lawyers, individuals), where registration of domain names is coupled with creating websites and developing online services.



Not only is the domain name industry in the region lacking in maturity, but so are related industries, according to the Egyptian registry. Examples include the relatively high cost of creating Arabic content, lack of appropriate mechanisms to protect intellectual property rights online, and low penetration and support for providing online payments or e-services.

Because of the immaturity of the market, the Egyptian registry points out that retail prices tend to be high, as economies of scale are not yet possible. Complex registration policies in the region, while often reflecting cultural and religious norms, tend to depress registration volumes. Examples of such policies include a requirement that applications for a domain name are backed by a letter (sent by mail or fax) on company letterhead, signed by a manager to support the application, residency requirements, requirements for domain names to reflect existing trademarks, and prohibitions on registrations of certain categories of name (eg geographical terms, tribal names, obscene or immoral words, names of apostles or prophets).

Not only do registration policies tend to be conservative, but management of character variants adds complexity. Key challenges, such as what language (or script) should be used to record the contact details of domain name holders, have not yet been resolved.

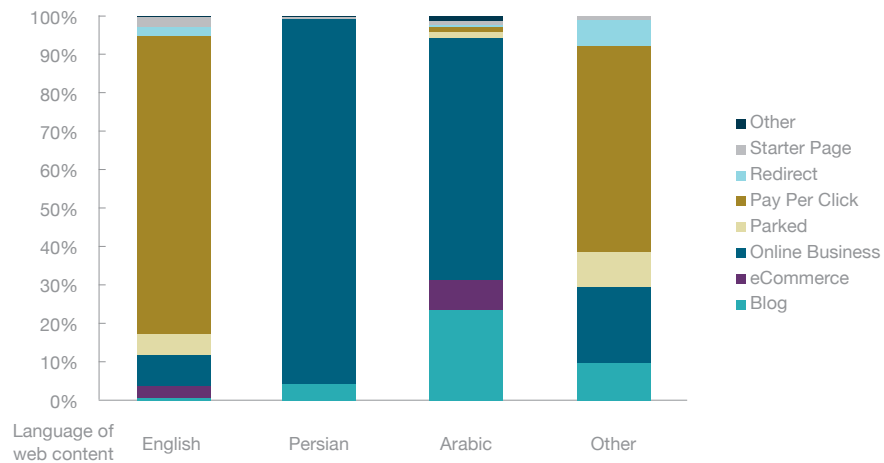
It has been observed already that lack of universal acceptance itself inhibits mass up-take (the IDN negative cycle).

The Egyptian registry points out a need for capacity building to develop local expertise, and develop the necessary technical, policy and business skills to build out the ecosystem of Internet industries at the local level. In the registry's view, exchanging best practices through forums such as APTLD, ICANN's Middle East DNS Forum can help, as can involving other relevant sectors in such knowledge-sharing.

7.4.2 How are Arabic domain names being used?

Verisign has analysed the type of web content associated with Arabic script second level domain names in .com and .net. The data sample was 27 000. The sample excludes those where there is too little text to determine the language of web content (21 000), and where the type of website cannot be determined (2 000), leaving 4,104 Arabic script .com and .net domains. The low usage rates exemplify the inherent complexity of combining right-to-left Arabic script second level domains, with left-to-right domain extensions (.com and .net).

Figure 38 – How are Arabic script .com and .net domain names being used?




Although English is the most popular language associated with .com and .net Arabic script IDNs, it is striking that nearly 90% of the sites are typical of low involvement web content – meaning that the domain name holder has invested little or no energy in content creation (pay per click, parking pages, starter pages, and redirects). Similarly, 70% of “Other” language websites (none of which are strongly associated with Arabic script), are used for low involvement content.

A very different picture emerges where the language of web content is Persian and Arabic. These have “high involvement” content (blog, online business, ecommerce), indicating that the domain name holder has invested energy and resources into content creation.

Nevertheless, the majority of Arabic script domain names in .com and .net which are associated with Persian and Arabic language content are used for online business purposes. This fits with general usage patterns:

- Individuals are tending to migrate away from domain names to social networks, but a minority create blog sites to publish their writing (some organisations use blog sites of course).
- Businesses are more likely to register and use domain names to create an online presence

Note that the data sample is small, and therefore generalisations may be of limited value.



However, in this small sample, Arabic script IDNs are more likely to be used for “high involvement” sites where the language is strongly associated with the IDN script (eg Persian or Arabic). Where the language of web content is not strongly associated with the IDN script (English, Other), the content is “low involvement”.

7.4.3 Strong correlation between local language content, country of hosting and script of IDN

We have seen that local language content is more likely to be hosted in country.

In the .com and .net sample of Arabic IDNs, 1132 Persian language websites were categorised as “online business”. Of these, 82% were hosted in Iran. Only 35% of Persian language blog sites were hosted in Iran. The reason for this difference is not clear. It may be that business websites tend to be more sophisticated, favouring a personal relationship between the business and web developer in the business’ own language or locality. In contrast many blog sites, if not exclusively for personal use, can now be set up in multiple languages, using popular templates (eg WordPress) without the involvement of a web developer.

The hosting country for Arabic language / Arabic script IDNs in the sample of .com and .net names were overwhelmingly out of region (98%). This goes against the correlation found in the ISOC study, but the data sample is small (215 domain names) so it is difficult to draw general conclusions from this.

7.4.4 A new approach in the region

A number of country code TLDs in the Arab States have changed management and strategy in recent years, introducing more liberal registration policies, accreditation of registrars, greater automation and lower retail prices. Examples include Qatar and the United Arab Emirates.

The United Arab Emirates registry manager, aeDA, explains that changes include a sustained outreach program, and regular contact with registrars to foster improvements in service quality.

The results have been impressive. The .ae ASCII TLD has grown to more than 110 000 registrations (a growth rate of 10% since September 2012⁶⁴), making it one of the largest TLDs in the region. Its experiences may be influencing the strategic direction of others – for example, the Omani TLD manager is currently considering a liberalisation strategy.

7.4.5 Other developments in the region

The Saudi registry has been working on universal acceptance issues related to IDNs, including IDN email. It has developed an extensive suite of tests which highlight the difficulties in using Arabic script domain names across browsers, applications, and email. It has also developed an email system which sends and receives fully Arabicised emails within a closed system (ie both sender and receiver have to be on the same system).

The Islamic Republic of Iran is preparing to launch its IDN ccTLD, ايران. The application was made to ICANN in 2009 as part of the IDN ccTLD Fast Track process. Although it successfully completed evaluation in 2010, obtaining the necessary government approvals delayed the implementation process until late 2013. The ايران. domain was launched in May / June 2014. The registry, IRNIC, has offered IDNs under .ir since 2006. When the ايران. TLD completed its evaluation stage in 2010, IRNIC stopped accepting new registrations of IDNs under .ir. At that point there were 6 000 IDNs, the highest number achieved in the region, despite the inherent difficulties in using domain names which combine right-to-left script with a left-to-right TLD (hybrid IDNs).

Qatar continues to report low uptake of IDNs. Despite excellent broadband penetration, a highly literate population and liberal ccTLD policies, Qatar has a high proportion of immigrant workers (85%⁶⁵), making English a popular language. The Qatar registry has also been active in research and development, and in advocacy for the IDN. It has developed and launched a mobile app for registering both .qa and قطر domains.


7.4.6 Task Force on Arabic Script Domain Names⁶⁶

The *Task Force on Arabic Script Internationalized Domain Names* is an initiative of the *Middle East Strategy Working Group* which focuses on technical issues and solutions related to Arabic script IDNs to promote their definition, secure deployment and ease of use for the community. ICANN has committed to fund and coordinate the work of the Task Force.

The Task Force currently consists of 26 members from 15 countries covering more than 10 languages of the Arabic Script. 20 members are new faces to Arabic Script IDNs. Two members of the group are linguistic experts in African Arabic Script languages.

The Task Force's work includes:

- developing rules for generating Arabic Script labels, both at the top level and second levels
- rules for recording user contact details for Arabic script domain names (these are currently required to be in Latin script), and
- issues relating to universal acceptance, character variants, associated software, security and training.



As with all work coordinated through ICANN, teleconference calls recordings and email archives can be found online⁶⁷. The Task Force is expected to conclude its work by the end of 2014.

The work of the Task Force is an essential building block in the path to universal resolution of Arabic Script domain names. Arabic script not only represents characters differently according to their position in a word, it also has zero-width-joiners, and many character variants. Unless these are handled in a coordinated fashion across different countries, Arabic script domain names will not work predictably across country borders. The work of the Task Force is likely also to influence the handling of Arabic script characters at the second level (usually the purview of individual domain registries), as many of the region's ccTLD managers are involved with the project.

FOCUS F

ICANN's new gTLD
programme

8 New gTLD IDNs

8.1 Overview

In January 2012, ICANN opened applications for new generic Top Level Domains (new gTLDs). The application process was crafted through ICANN's policy development process over a seven year period, beginning in 2005. More than 1,900 applications were received. The evaluation process lasted several months, and the first new gTLDs started to come to market from late 2013.

At the time of writing (May 2014), 252 new gTLDs have launched of which 23 are IDN new gTLDs (9%)⁶⁸. The total number of new gTLD domain name registrations is 715 000, of which 54 000 (8%) are in IDN new gTLDs. While the percentage of IDNs in the new gTLD space is low (8%), it is higher than the percentage of IDNs in all domain name registrations (2%).

8.2 Meeting unmet needs and supporting multilingualism

Currently, only 2% of all domain name registrations. The majority of existing domain names are in ASCII. One of the objectives of the new gTLD programme was to meet unmet needs for the global population. An obvious gap is linguistic diversity in the domain name system.

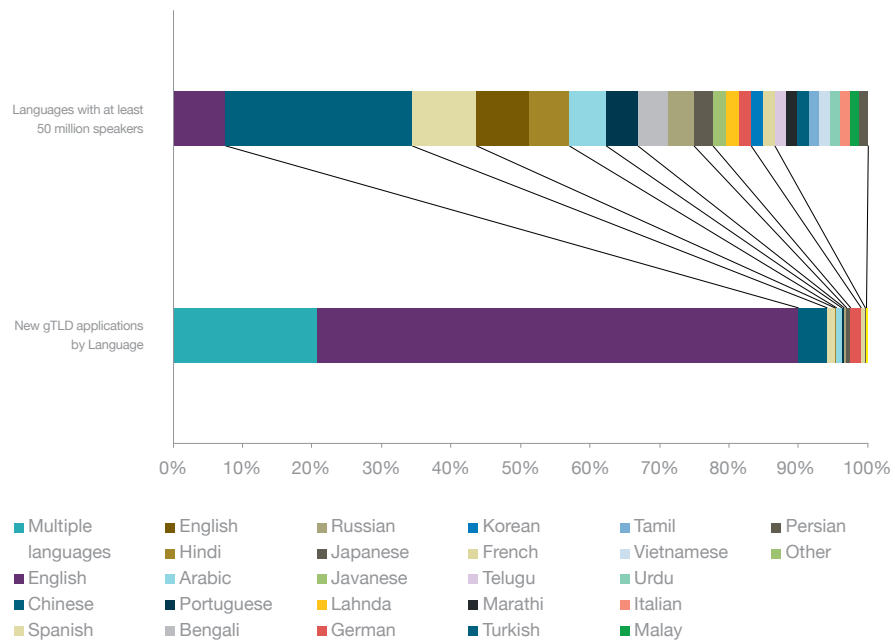
8.2.1 New gTLD applications – do they represent the world's most popular languages?

Sadly, the new gTLD applications do nothing to enhance online linguistic diversity. Analysis of the gTLD applications, whether by script or language, confirms the primacy of the English language in the domain name system. Figure 39 compares the language of new gTLDs with the world's top languages.

We analysed the language of 1165 new gTLDs⁶⁹. Sometimes, it was not possible to identify the language of the gTLD string from the application, and in 244 applications the string was meaningful across multiple languages (typically English and French, English and German, Spanish and Italian), for example .app, .golf, .poker, .salon. Of the strings identified as "multiple languages", over 220 are meaningful in English. Therefore, the overall picture remains persistently English language dominated, with up to 90% of new gTLD strings being meaningful in English.



Figure 39 – Language of new gTLD applications compared with the world’s top languages



Up to 90% of new gTLD strings are meaningful in English”

A similar picture emerges when viewed by script. 94% of new gTLD applications are in Latin script. Of all other scripts, only Han (Chinese and Japanese) has over 1%. Only 3 new gTLD applications – for transliterations of .com, .net and .org – are in Devanagari, the writing system for the world’s fourth most popular language (Hindi, 260 million speakers⁷⁰). The scripts necessary for writing Bengali, Javanese, Lahnda, Telugu, which are among the world’s most popular 20 languages are not represented in the new gTLD applications at all.

8.2.2 Early market performance of new gTLDs (including IDNs)

Generally, the level of registrations in the new gTLDs as a whole has been disappointing, compared with projections⁷¹, and the levels of venture capital raised⁷².

At the time of writing (May 2014), there are two IDNs in the top 20 new gTLDs (and 7 in the top 100). A familiar long tail pattern is apparent in the new gTLD registration volumes.

The IDN new gTLDs with the highest volumes of registrations to date (May 2014) are .在线 (“ .online” in Chinese) and .中文网 (“ .website” in Chinese) (see case study at 8.2.5).

In April 2014, the first two Cyrillic domains, .CAЙT and .ОНЛАЙН were offered for general registrations. ОНЛАЙН achieved 1400 registrations, and CAЙT 667 registrations by May 2014. Again, the volume of registrations is low compared to .рр which achieved over 600 000 registrations in its first month (November 2010).

Before the new gTLD program, and despite the success of some IDN ccTLDs, the domain name system was dominated by English language strings and registrations. An unmet need was to introduce multilingualism into the domain name system. Both on an analysis of language of new gTLD strings, and early new gTLD registrations, the program appears to be failing to meet the unmet need of a multilingual domain name system.


Figure 40 – Arabic script new gTLD applications⁷³

.موقع .شبكة .كيوتل .موزايك .
.كاتوليك .عرب .بيتك .العليان
.اتصالات .بازار .ابوظبي
.موبايلي .همره .كوم .ارامكو

8.2.3 Case study on شبكة.

شبكة. (meaning “Internet” in Arabic) was the first new gTLD to sign its contract with ICANN (July 2013), and the first to come to market (November 2013). The vision of شبكة is to provide a “truly Arabic Internet experience from start to finish⁷⁴. The DotShabaka registry notes the boom in the growth of Arabic online content, and the fact that 380 million Arabic speakers live in the Arab States.

8 domains were registered in the sunrise period (November 2013-February 2014). This low figure is in line with the low registration numbers during sunrise periods across all new gTLDs. Further, according to the DotShabaka registry, given the complexities of the different ICANN-mandated opening phases (sunrise, landrush), they did not attempt



to market شبكة. until the general availability phase began. At that point (February 2014), over 1 000 registrations were made between 29 January and 6 February 2014⁷⁵. By early May 2014, registrations under شبكة had risen to 1 889.

75% of شبكة registrations are with the registrar 101Domains.com, which specialises in multilingual domain names. Many of the remaining 25% are registered with brand protection registrars (such as Com Laude (8%), IP Mirror (2%), CSC (2%), Marcaria (2%), Mark Monitor (1%)), suggesting that these are likely to be defensive registrations rather than new registrants.

The DotShabaka registry reported in May 2014 that only 1% of its domain names were in use. The registry cites two factors as hampering uptake: first, lack of universal acceptance for Arabic domain names; and second a lack of end user awareness that domain names exist in other scripts. Our analysis of 1679 domains (in March 2014) showed that nearly 1 100 had active name servers – an essential prerequisite to active use. However, of those with active name servers, nearly 1 000 had “insufficient characters” to analyse the language of web content. We found Arabic language in 31 شبكة websites, English in 78, and French in 2. This analysis, which took place just one month after the domain launched, should not be taken as a general indication of future usage.

8.2.4 Case study on みんな

みんな means “everyone” in Japanese, and is operated by Google.

It achieved 3 registrations during its sunrise period (December 2013-March 2014). Google launched the みんな domain by offering free domains to the first 5 000 people to submit a website idea⁷⁶ via a campaign website, which featured an animated video in Japanese language⁷⁷. From March 2014, みんな quickly achieved 2 500 registrations. Since then, growth has levelled out, with the domain reaching 3 500 by early May 2014⁷⁸.

The Japanese registrar, Interlink Co. Ltd⁷⁹ has 82% of the みんな registrations, emphasising the link we have found elsewhere between local registrars and local language. 101 domains, a more generalist IDN registrar has 8%. Brand protection registrars have less market share in みんな than we observed in شبكة.

Our analysis of 3478 みんな domains (in May 2014) showed that only 1 000 had active nameservers. Of those with active nameservers, 258 had “insufficient characters” to analyse the language of web content. We found Japanese language in 97 みんな websites, English in 428, and other languages in 15 instances. This analysis, which took place less than two months after the domain launch, should not be taken as a general indication of future usage.

8.2.5 Case study on 中文网 (.Chinese) and 在线 (.Online)

中文网 means “Chinese” and 在线 means “Online” in Chinese. Both are operated by the TLD Registry Limited. Marketing materials for the two Chinese TLDs emphasise the intuitive nature of the names for Chinese users, and the search engine optimisation benefit of domain names that exactly match common search terms.

中文网 and 在线 achieved 310 and 226 registrations in their sunrise periods respectively (January-March 2014). There followed a month-long landrush period, during which registrations grew to 8 000 and 10 000 respectively.

From late April 2014, 中文网 and 在线 quickly grew to 14 000 and 30 000 registrations early May 2014⁸⁰.

In 中文网, the Chinese Government has 55% of registrations (acting as a registrar), Xin Net Technology 5%, and Xiamen Nawang Technology 1%. The German registrar 1API (Hexonet) has 36% of registrations.

In 在线, the German registrar 1API GmbH (Hexonet) again performs strongly with 64% of registrations. Nearly 30% of 在线 registrations are with Chinese registrars (Chinese Government 25%, Xin Net Technology 4%). 101 domains has more than 3%.

We are unaware of any particular connection between Hexonet and the 在线 domain. The registrar experienced a spike of 25 000 in its new gTLD registrations coinciding with the date of general availability of 中文网 and 在线 (late April 2014)⁸¹.

Brand protection registrars have less market share in 中文网 and 在线 than we observed in شبكة.

As both 中文网 and 在线 launched in late April 2014, it was too early to make a meaningful analysis of usage rates at the time of writing.

FOCUS G

IDN facts and figures

9 IDN facts and figures

9.1 Total IDN registrations (IDN ccTLDs, IDN gTLDs and IDNs at second level)

There are **6 million IDNs registered** (Dec 2013). IDNs represent just 2% of the total number of domain names registered in the world.

Our data sample of **270 million** domain names comprises over 99% of the world's registered domains⁸².

9.2 Deployment of IDNs by domain name registry

According to our data, there are now 74 IDN deployments⁸³, 50 at the second level (eg παράδειγμα.eu) and 24 IDN TLDs (eg 例子.中国). Until 2013, only ccTLDs had implemented IDN TLDs. In 2012, ICANN received over 1 900 applications for new gTLDs, including IDN TLDs. In late 2013, the first of approximately 100 IDN new gTLDs were launched, and the remainder will be launched over the next 2 years or so. India will be launching several IDN ccTLDs to support official languages (2014), and we are aware of 10 TLDs actively considering IDN deployment at the second level.

During the years 2000-2009, when it was impossible to register full IDNs, there was no choice but to launch IDNs at the second level (or not at all). Registries from the Asia Pacific region for example in Republic of Korea, China, Hong Kong SAR China, Taiwan of China, Japan, adopted IDNs at the second level during those years, and from the years 2009-2013 have also launched full IDN ccTLDs.

The tendency within the Middle East and North Africa (with the exception of Islamic Republic of Iran) was to wait for full IDNs to become available, in order to avoid the complexity of handling a mix of right to left Arabic script and right to left Latin script endings. Registries from within the region have been adopting full IDNs in the years 2009-2013.

In the coming years, we anticipate that the number of full IDNs may overtake the number of second level IDNs. It will be interesting to see whether registries in non-Latin script countries will continue to maintain both second level and full IDN offerings. For example, Islamic Republic of Iran, an early adopter of IDNs at the second level ceased new registrations while it was awaiting approval of the ايران. On the launch of ايران., the registry moved across existing second level IDNs from .ir to .ایران.

Figure 41 – IDN launches per year (second level v IDN TLD)

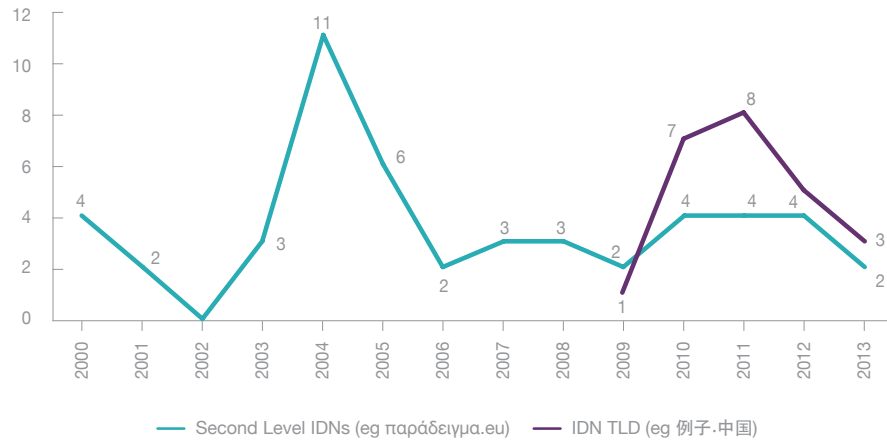
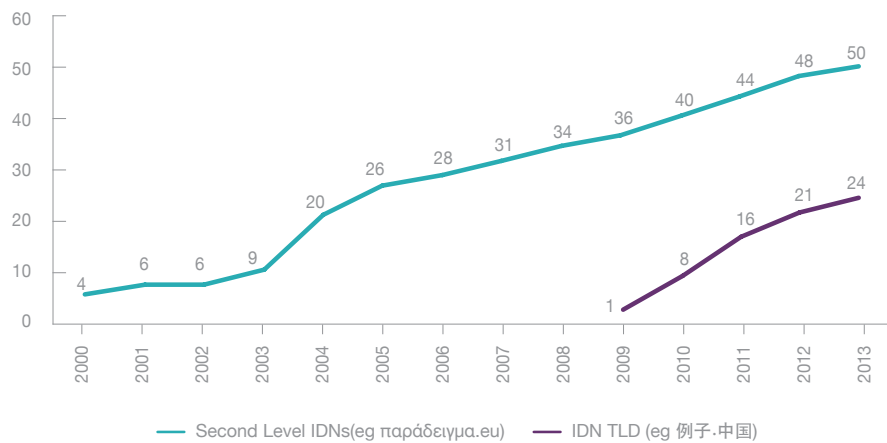


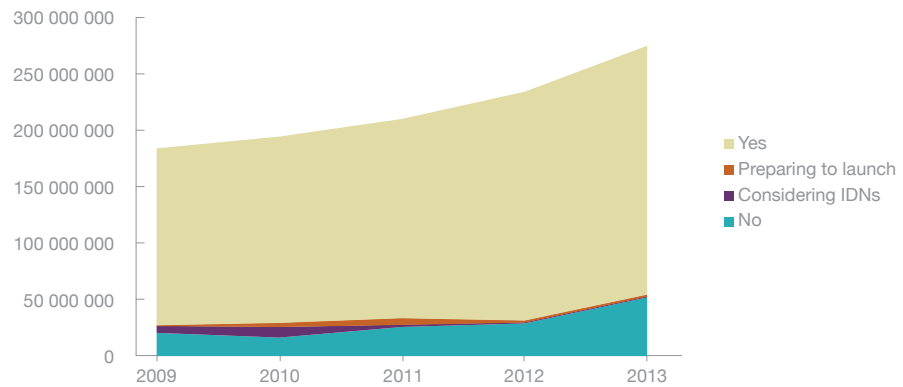
Figure 42 – IDN launches per year (second level v IDN TLD), cumulative



Reviewing the number of IDN launches per year since 2000 (figure 41), there was a peak in launches at the second level in the middle of the last decade. More recently, the trend has been towards full IDNs. The exception is for Latin script IDNs, where the ASCII ending is meaningful. For example, during 2013 the Belgian and Canadian ccTLDs launched IDNs at the second level to support French speakers in their communities. Even in countries where the local language uses non-Latin scripts, the ASCII domain extensions persist, perhaps for historical reasons, having been the extension since the Internet was brought to these countries.



**Figure 43 – Does your registry allow IDNs?
By domains under management 2009-2013**

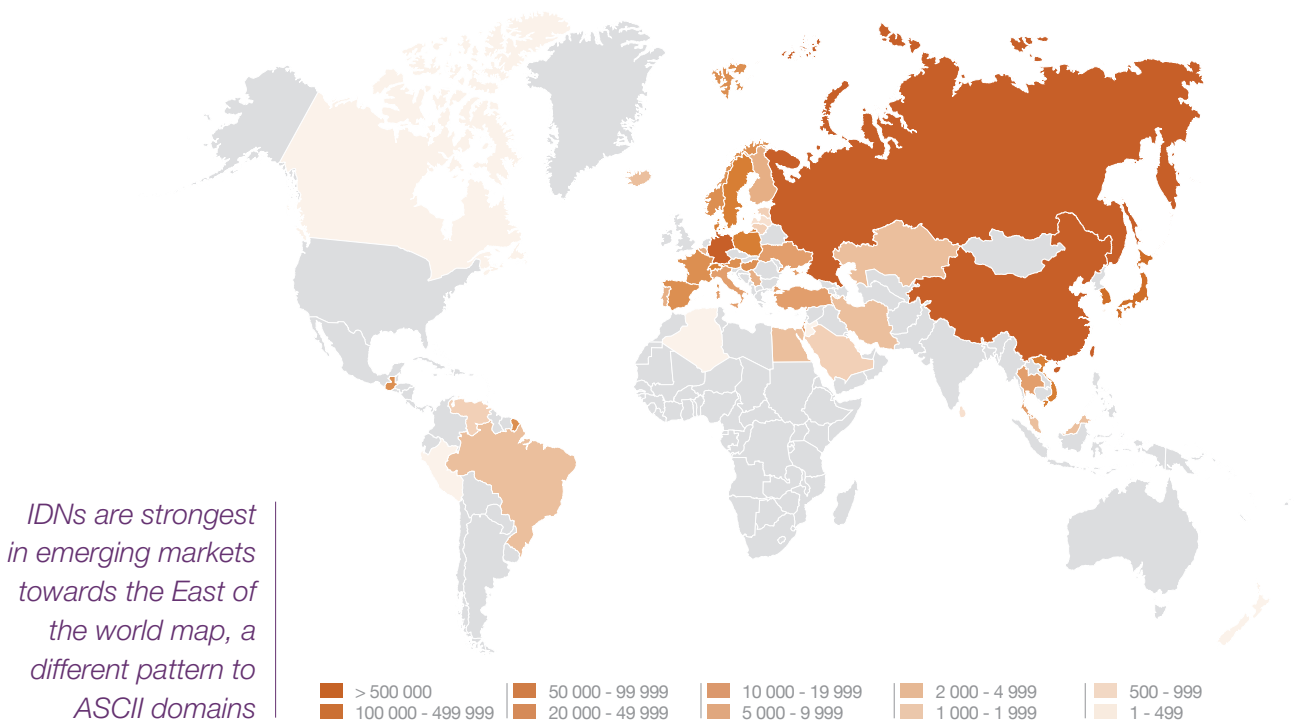


When viewed by the number of domain names under management over time (figure 43), it seems that the market is settling. The majority of domains under management (80%) are within registries that support IDNs. As we noted in the 2012 report, the process of IDN deployment takes 2-3 years from the point where it is first considered. The numbers in transitional stages (“considering IDNs”, “preparing to launch”) are decreasing year on year, whether considered by number of registries, or number of domains managed. This suggests that established registries are either completing their IDN implementation cycles, are deciding against or postponing implementation.

9.3 IDNs perform strongly in Russian Federation, Asia & Pacific

This year, we have created heat maps showing geographical distribution of IDNs and all domain names. The comparison includes ccTLDs only, and excludes gTLDs as they are not geographical in nature.

**Figure 44 – The IDN World Map 2013 –
Excludes gTLDs. Includes ccTLDs (both full IDN, and hybrid IDN)**



IDN ccTLDs are relatively recent, having been deployed from 2009 onwards. Some countries which have long been planning IDN implementation have not yet come to market, for example India. Registrations of IDNs are most intense towards the East of the map, for example, Russian Federation, Viet Nam and China, with strong showings also from Republic of Korea and Japan.

Within Europe, IDNs at the second level are most popular in countries strongly associated with Latin-script diacritics and accents, for example Germany, Sweden, Norway, France and Spain.

Within the Arab States and Latin America and Canada, IDN registration levels are generally lower.

In contrast, countries like the UK, US, Netherlands and Australia have not implemented IDNs – reflecting that their populations are well served by ASCII domain names.

We do not have data from much of Africa, but hope in future years to form a stronger partnership with the African ccTLD organisation and individual registries.

9.3.1 IDNs by region 2009-2013 (IDN ccTLDs and IDNs at second level)

Figure 45 includes all IDNs (both second level and IDN TLDs) by UNESCO geographic region (with Islamic Republic of Iran included with other Arabic script IDN registries within “Arab States”) plus gTLD IDN registrations⁸⁴.

An analysis of the IDN registration figures (both IDN ccTLDs, and IDNs at the second level) in figure 45 shows that IDNs in the Asia and the Pacific region has had the greatest net growth between 2011-2013. Much of this is attributable to the growth of .vn IDN domains. Following a change in policy in 2011, .vn IDNs are currently given away for free. Last year, we anticipated that .vn IDNs would reduce dramatically on renewal, given the low usage rates (then 10%). This has not occurred, in fact .vn IDNs have grown by 14% in 2013. The .vn figures continue to mask a net reduction in IDNs in the Asia and the Pacific region between 2011-2013, with Republic Korea in particular suffering a 34% reduction in its IDN ccTLD, .한국.

While IDN registration numbers in gTLDs and the European and North American region have remained relatively steady through 2011-2013, these two categories have seen the greatest erosion of market share in the past five years (from a combined 84% to the current 58%), as registrations in Asia and the Pacific have grown.

For 2013, we have data for Latin American thanks to the cooperation of the regional TLD organisation, LACTLD. While the region’s share of the global market is less than 1%, the distribution of IDNs is shown in figure 46. As we have no data prior to 2013, we have shown two data points during 2013 (June and December). Even with comparatively low numbers, there is positive growth (4%) in the six month period, with Venezuela and Brazil having the largest IDN registrations in the region.

Figure 45 – IDNs by geographical region (both second level and IDN TLD) 2009-2013

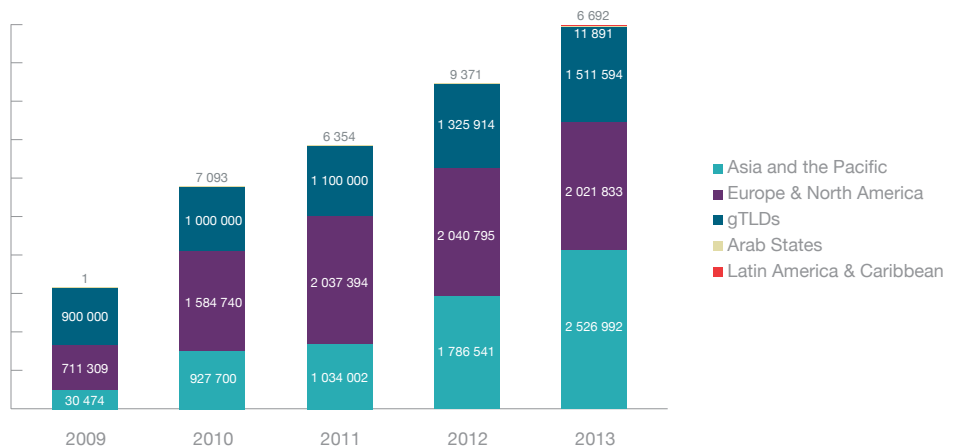
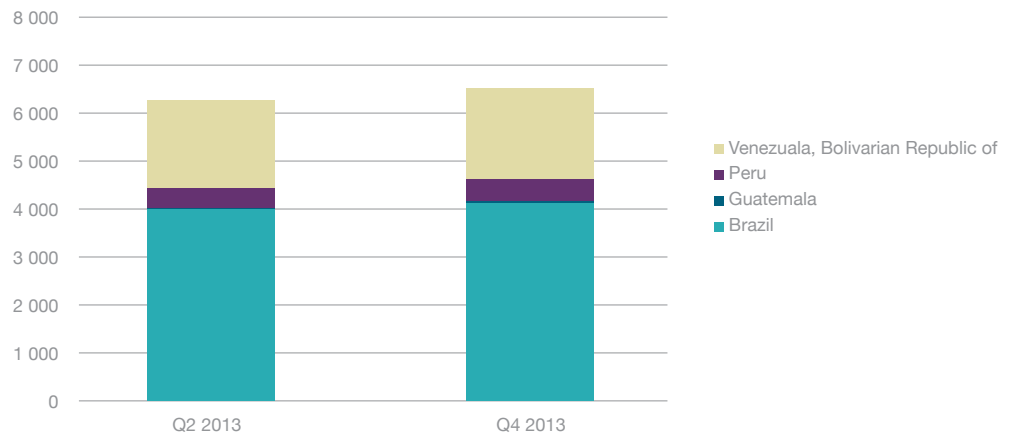
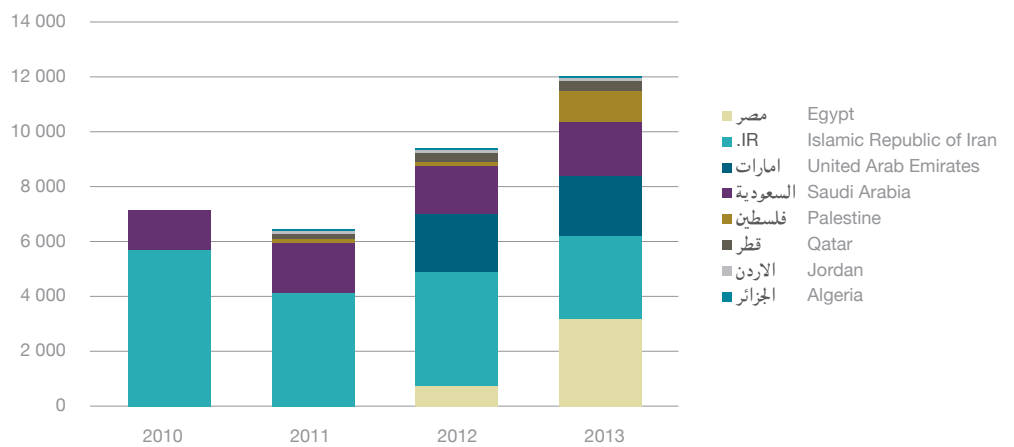


Figure 46 – IDNs in Latin America and the Caribbean region 2013
(source: LACNIC)



Across the Arab States and Islamic Republic of Iran, countries which have deployed Arabic script IDNs have less than 1% global market share of IDNs. However, registrations of IDNs have increased year on year since 2011 despite reductions in the then market leader .ir, with the Egyptian IDN coming through strongly in 2013 (see figure 47).

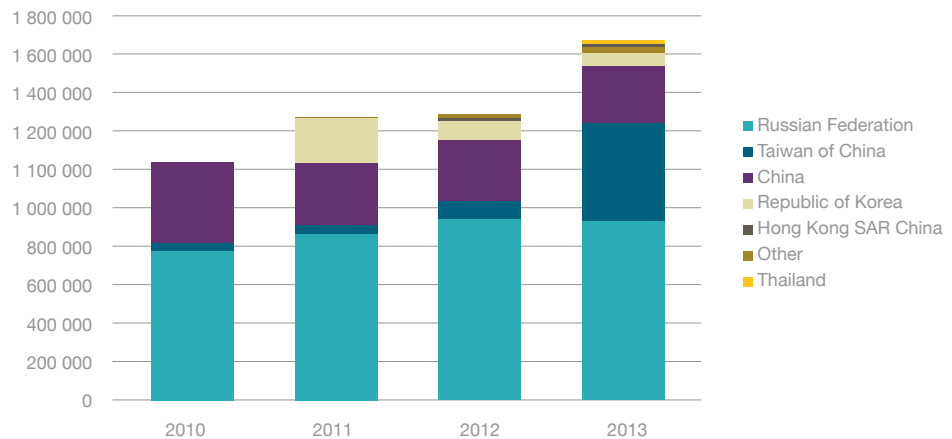
Figure 47 – Arabic script IDNs 2010-2013



9.3.2 IDN ccTLDs

IDNs occur at the second level eg παράδειγμα.eu or as an IDN ccTLD eg 例子.中国. The charts above show figures for all IDNs in our data sample, comprising both IDN registrations at the second level and IDN ccTLDs. Figure 48 contains only the IDN ccTLDs.

Figure 48 – IDN ccTLDs by year (2010-2013)



There has been strong growth in IDN ccTLDs between 2012-2013, with Taiwan of China showing high growth thanks to a bundling policy, whereby new registrations are mapped across both ASCII, traditional and simplified Chinese TLDs. Republic of Korea continues to contract, whereas China has remained steady after reductions between 2010-2012.

9.3.3 How does IDN growth rate compare with the growth of ASCII domain names?

Between 2009-2013, the number of IDNs has tripled (figure 49), a growth of 215%⁸⁵. Note that the numbers are lower than overall domain name registrations, so high percentages can result from small changes in numbers.

While growth rates are encouraging, the number of IDNs remains low in comparison to overall domain name registrations, only 2% of the total. Compared with the number of speakers of languages which rely on non-Latin scripts offline, IDNs continue to under-perform compared to their potential.

Between 2009-2013, IDN growth rate has been 215%

Figure 49 – Total IDNs growth 2009-2013

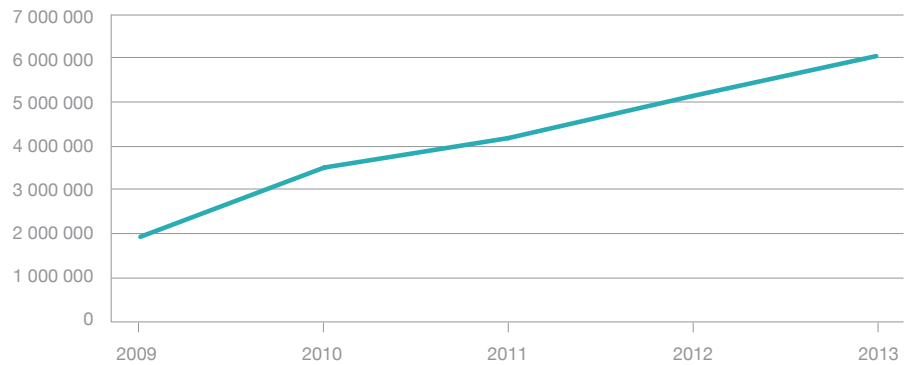
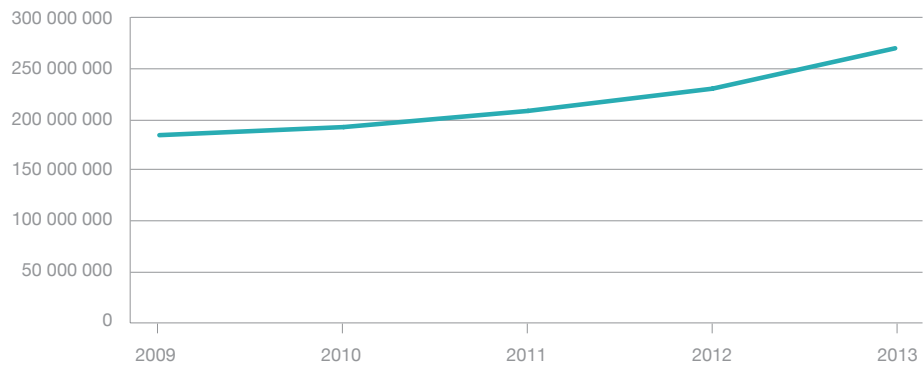


Figure 50 – Total domains growth 2009-2013



Between 2009-2013, the total number of registered domain names has grown by 41% (figure 50). While the growth rate is much less steep than that of IDNs, the overall numbers are far higher (192 million in 2009 to 271 million in 2013).

Figure 51 shows a comparison of the growth rates per year 2009-2013. The 84% annual growth for IDN (2009-2010) coincides with the launch of pφ, and many other IDNs (including .eu). Since then, annual growth rates for IDNs have averaged 14%, higher than that of total domain names (10% for 2010-2013).

Is this a sign that the IDN market is maturing? If we drill down to the level of registries which operate both IDNs and ASCII TLDs, we see that the IDN growth rates are much more volatile than that of ASCII domains (figure 51).



Figure 51 – IDN and total domain annual growth rate compared 2009-2013

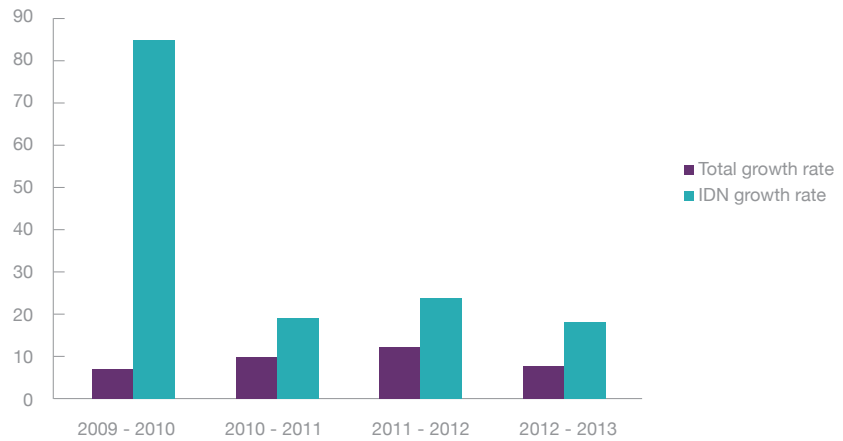
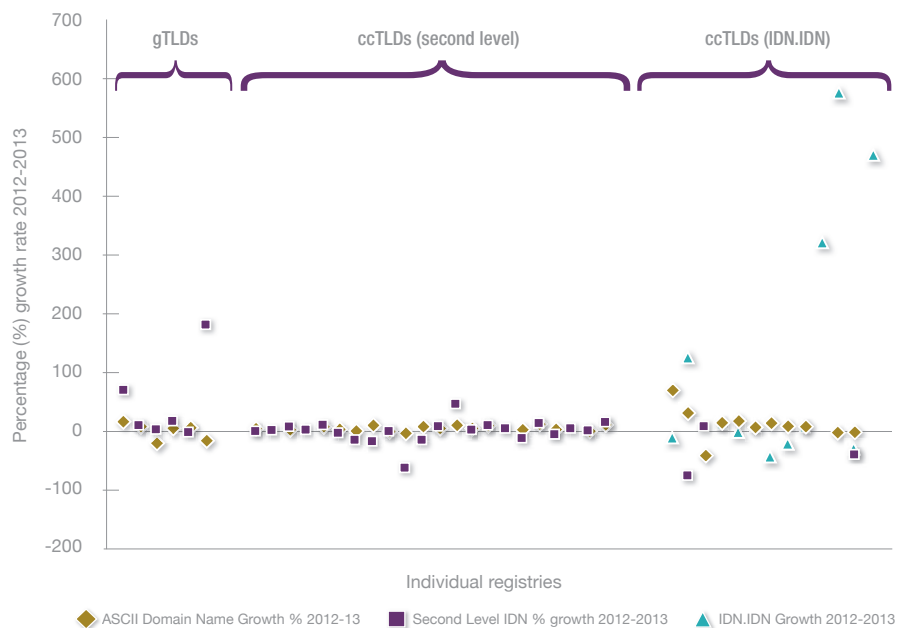



Figure 52 compares growth rate of a selection of 41 registries where we have data covering both growth of ASCII domain names and IDNs (second level IDNs and IDN ccTLDs) for 2012 and 2013.

Figure 52 – Comparison of annual percentage growth rates in sample of ASCII domains, second level IDNs, IDN.IDN 2012-2013





The graph shows separately gTLDs, ccTLDs (registering IDNs at the second level) and ccTLDs offering IDN TLDs. Some ccTLDs, eg Republic of Korea, Hong Kong SAR China, offer IDNs both at the second level under the ASCII country code (eg .kr, .hk) and full IDN TLDs.

The results show:

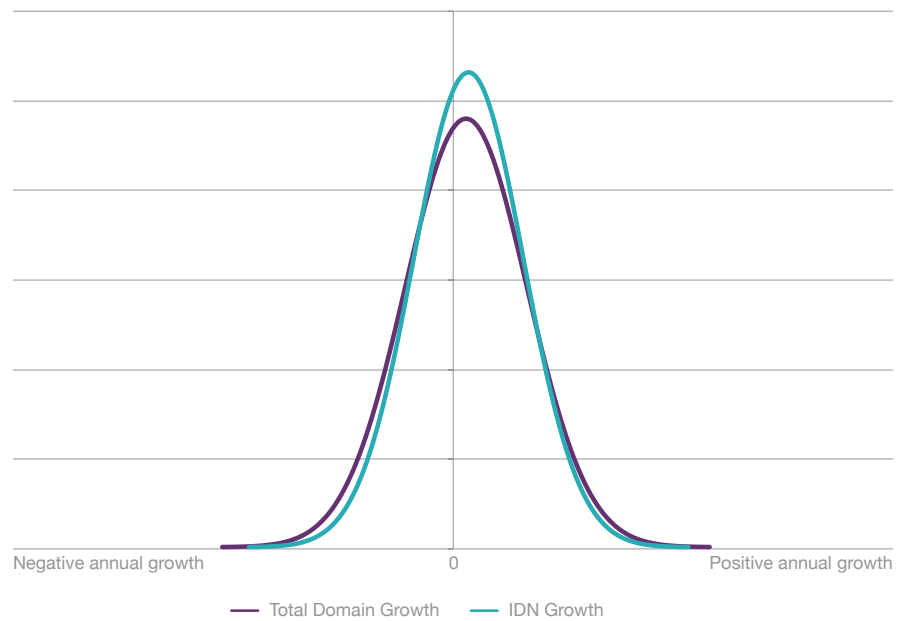
- For gTLDs, ASCII growth rate averages -2.5%, thanks to drops in both .info and .asia. IDN growth rate averages +46%, thanks to high growth in .biz (+68%) and .asia (+181%) over the past year.
- For ccTLDs which offer IDNs at the second level, there is generally less volatility than is observed in gTLDs and IDN ccTLD registries. The average ASCII growth rate in the year is 4% and there is a negative growth in IDNs at the second level (-5% average). Second level IDNs in several European ccTLDs (eg .it, .eu, .fr and .pl) have seen double-digit negative growth in the past 12 months. Further, some large drops also skew the picture, eg:
 - .hk second level IDNs have dropped by -77% in the year. At the same time, the IDN ccTLD for Hong Kong SAR China has grown by +125%;
 - .il has a negative growth of -62%, an understandable drop after a healthy initial uptake (the “landrush renewal” phenomenon), and the difficulties of combining right-to-left Hebrew script with left-to-right Latin script TLD ending “.il”
- IDN ccTLD domains continue to show volatility, and because some of the numbers are low, the percentage growth rates can appear very high. For example, فلسطين (Palestine) has grown +581% from 163 domains to 1110 in the period).
 - Taiwan of China has grown by over +400% in part because of its bundling policy
 - Egypt has grown by +325% after opening for general registrations in early 2013.
 - Some of the larger IDN ccTLDs (Republic of Korea and China) have experienced negative growth in the past 12 months (-34% and -10% respectively)
 - Serbia has experienced negative growth, approximately 1 year after launch (-44%). This may be an example of the landrush renewals phenomenon

Across the entire sample, the average (mean) growth of the ASCII domains in our sample was 3% (2012-2013) with a standard deviation of 18. The average growth rate for IDNs at the second level was 5% (2012-2013) with a standard deviation of 114. The IDN ccTLD average growth rate was 116%, with a standard deviation of 204, indicating a high volatility. The median for the ASCII domains in the sample is 5%, and for IDNs is 2% (second level) and 7.5% (IDN ccTLD).

Using scaling to allow for comparison, we have plotted the normal distribution of both IDNs and total domains on the same growth (figure 53). Both describe a classic bell curve, with the median growth being positive. The positive messages, that IDNs have healthy growth rates, in line with total domains, is sometimes getting lost in the low overall IDN figures.



Figure 53 – Bell Charts 2013



Using percentages as the measure of growth introduces some distortion. For example, .com added 6 million domains (2012-2013), a growth rate of 5% from 108 million to 114 million. Meanwhile, by losing 7 000 domains, coop shrunk by -48% (15 000 to 8 000). The distortion is particularly felt in the IDN TLDs, several of which still have under 1 000 domains.


9.4 Conclusions

2013 has seen continued growth in IDN registrations, both at the second level and IDN ccTLDs. There are now 6 million IDNs in the world, and usage rates are steadily growing. 2013 also brought to market the first IDN new gTLDs.

Where IDNs are in use, they are accurate predictors of the language of web content. People are choosing to signal the presence of local language content with IDNs, and IDNs are also more likely to be hosted in-country.

The field of cognitive science helps us to understand the complexity of the processes through which the brain understands written language. Factors such as ease of recall, context and understanding strongly indicate the need for the Internet's navigation system to reflect local languages.





2013 has brought progress in improving universal acceptance of IDNs, such as the handling of URLs in social media and search, and the recent announcement that Google is supporting internationalised addresses in Gmail. Much remains to be done, however, and universal acceptance needs to be considered throughout the Internet “stack” (from infrastructure, through addressing, to applications).

The Arab States have seen a burgeoning in growth of Arabic language content in recent years. There is enthusiastic use of social networks, with user-generated Arabic content, throughout the region (particularly in Egypt and Saudi Arabia). Yet, registration of domain names (whether ASCII or IDN) remains low across the region. We believe that multiple factors are at play, highlighted in the IDN readiness matrix. Colleagues from the region also point out the need for capacity building, and the relative gaps in related industries, which lead to relatively high cost of creating Arabic content, lack of appropriate mechanisms to protect intellectual property rights online, and low penetration and support for providing online payments or e-services.

New gTLDs promised to fulfil unmet needs in the domain name system. One obvious need is for enhanced linguistic diversity. With 90% of new gTLD strings (eg .photography) either in English language, or understandable in English, that opportunity has been lost – for now.

Growth rates of IDNs remain positive, and the correlation between IDNs, language and country of hosting illustrate the potential of IDNs to provide a gateway to local language content. This is a vital aspect of maintaining a single, diversified Internet in the future. Already, half of today’s Internet users are from Asia. The growth potential for Internet penetration is in Asia and the Pacific, Africa, and Latin America where English is not the primary language. For IDNs to fulfil their potential, multiple actors need to make changes to hasten universal acceptance, so that IDN can be used seamlessly in every environment where an ASCII domain name would be used.



APPENDICES



APPENDIX 1


What are Internationalised Domain Names, and why are they important?

Domain names, the Internet’s addressing system, work because they are interoperable and resolve uniquely. This means that any user connected to the Internet, anywhere in the world, can get to the same destination by typing in a domain name (as part of a web- or email address). The plan to internationalise the character sets supported within the Domain Name System is almost as old as the Internet itself. However, technical constraints and the overriding priority of interoperability resulted in a restricted character set within the Domain Name System: ASCII¹ a to z, 0 to 9 and the hyphen².

Technical standards to internationalise domain names were developed from the mid-1990s. The solution retains the domain name system’s restricted character set, and transliterates every other character into it. Each series of non-ASCII characters is transliterated into a string of ASCII characters prefixed with xn-- , called Punycode. Punycode domain names are meaningless to humans, but meaningful to machines that resolve domain names – name servers. Thus, humans see the meaningful, transliterated characters when they navigate the Internet, whilst the underlying technical resolution of domain names remains unchanged.

Figure 1 – Internationalised Domain Names explained

	IDN second level	ASCII top-level domain (TLD)
human readable (UTF8):	παράδειγμα.eu	Greek script domain name (hybrid)
machine readable (punycode):	xn--hxajbheg2az3al.eu	The same domain name in punycode
	IDN second level	IDN top-level domain (IDN ccTLD)
human readable (UTF8):	образец.рф	Cyrillic script domain name (full IDN)
machine readable (punycode):	xn--80abnh6an9b.xn--p1ai	The same domain name in punycode



Implementation of IDNs began in 2000 at the second level (under .com and .net) and 2001 (.jp). In the ten years that followed, several ccTLDs deployed IDNs, primarily supporting local language character sets. Some experimented with other strategies for internationalising domain names, but the IDN technology proved the most successful. Following pressure from the ccTLD community, ICANN introduced a fast track process to create IDN ccTLDs in 2007-2008. From 2010, IDNs became available at the top level having completed the specific process set by ICANN (for example, السعودية for Saudi Arabia, рф for the Russian Federation)³.

IDNs are technically complex to implement. Many challenges remain, including (at a technical level) how to handle variant characters⁴, which are prevalent in Arabic and Chinese scripts. Another challenge is the user-experience, eg consistent representation in browsers and emails.

Despite the technical challenges, IDNs are viewed by many as a catalyst and a necessary first step to achieving a multilingual Internet. According to UNESCO, in 2008 only 12 languages accounted for 98% of Internet web pages; English, with 72% of web pages, was the dominant language online⁵. Recent reports indicate that other languages are growing rapidly online. For example, by 2010, only 20% of Wikipedia articles were in English⁶. Supporters of IDN believe that enabling users to navigate the Internet in their native language is bound to enhance the linguistic diversity of the online population, and that IDNs are strongly linked to local content.

While this study focuses on the web, it should be noted that other applications also require internationalisation, eg email, file transfer protocol, etc.

1 IDN timeline

For more than a decade, hybrid Internationalised Domain Names have been available at the second level with ASCII Top Level Domains (for example, παράδειγμα.eu in figure 1). This situation was only satisfactory for Latin-based scripts used by most European languages, where the IDN element would commonly reflect accents, or other diacritical marks on Latin characters. For speakers of languages not based on Latin scripts (for example, Chinese, Arabic), the hybrid IDN/ASCII domains were unsatisfactory. Right-to-left scripts, such as Arabic and Hebrew created bi-directional domain names when combined with left-to-right TLD extensions, requiring users to have a familiarity with both their own language, and Latin scripts in order to navigate the Internet. As explained in the report IDNs State of Play 2011, bi-directional domain names not only require Internet users to change script when typing in a single web address, but also potentially confuse the strict hierarchy of the Domain Name System. Industry experts describe bi-directional domains as “barely usable”⁷.

Figure 2 – examples of hybrid and bi-directional IDN domain names (Japanese, Arabic, Hebrew).



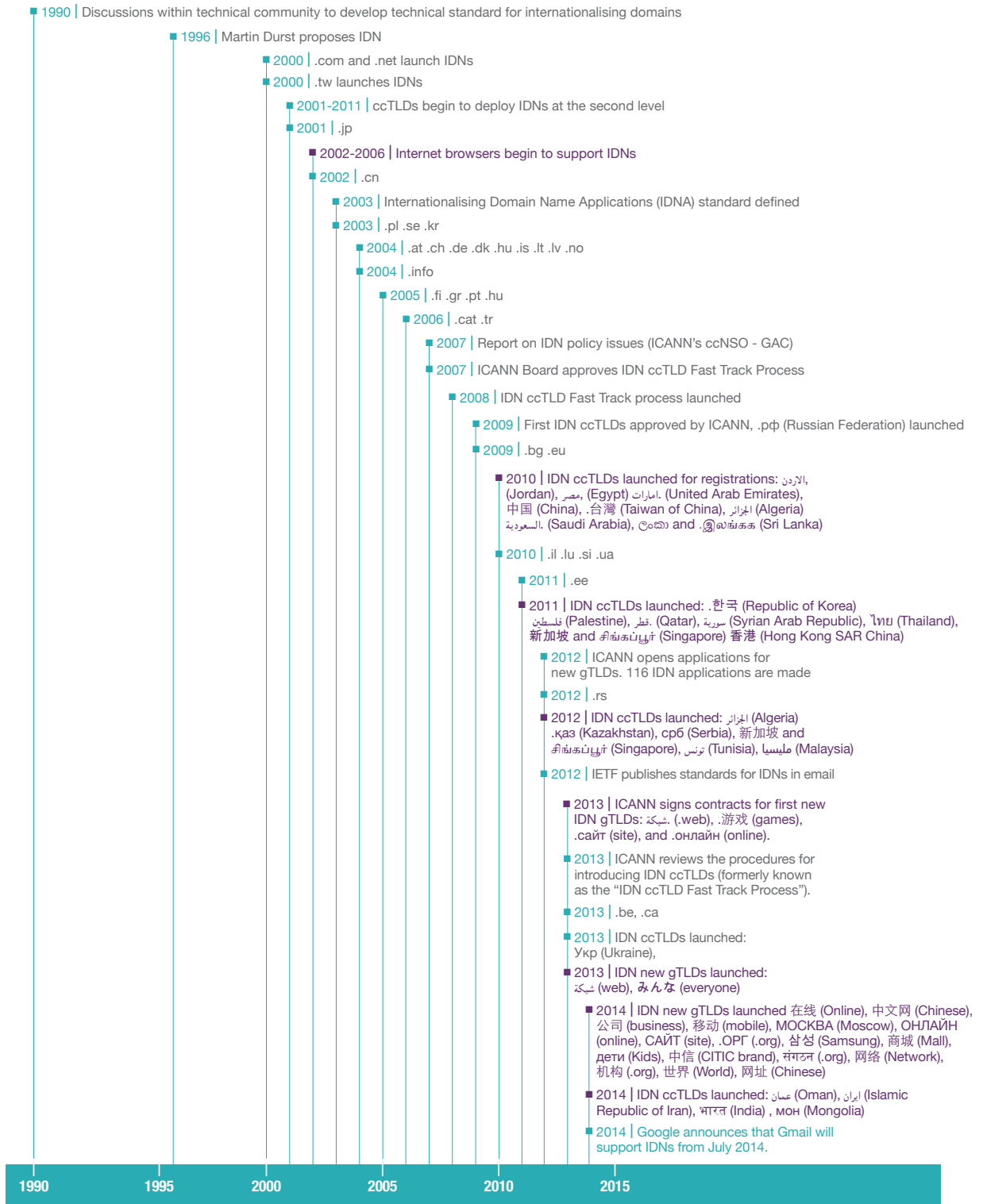
Internet governance discussions from 2006 onwards highlighted the lack of IDNs in the root domain zone (which would enable full IDN domain names including at the top level) as a key building block towards the goal of a multilingual Internet⁸. From 2005, there was increasing pressure on ICANN, the global coordinator of Internet domain names, to implement IDNs in the root zone.

In the meantime, some countries created their own work-arounds. For example, China and the Republic of Korea developed keyword searches at the domain name servers for .cn and .kr. For those searching for domains within the country, the keyword system resolves the domain without the user having to type the Latin-script domain ending (TLD). In China and Egypt, browser add-ons were developed to translate a domain into another name that would be looked up on national servers, to enable Internet users to enter local character strings into browsers. However, this solution relied on users downloading a plug-in, which was not compatible with every browser. These efforts indicate the importance that policy makers and technologists have placed on internationalising domain names, and that IDNs emerged as the superior technology amongst a number of alternatives.

In 2009, the ICANN Board approved a fast track process for IDN ccTLDs, describing the programme as a “top priority”⁹. By April 2011, 17 IDN ccTLDs had been launched. Since then, there has been a steady expansion of the number of IDN.IDN registries launched, including .한국 (Republic of Korea), قطر (Qatar), فلسطين (Palestine), الجزائر (Algeria), .香港 (Hong Kong SAR China), سورية (Syrian Arab Republic), .қаз (Kazakhstan), срб (Serbia), 新加坡 and சிங்கப்பூர் (Singapore).

In mid 2013, ICANN signed its first contracts for new gTLDs: .شبكة (.web), .游戏 (games), .сайт (site), and .онлайн (online). The new gTLDs started to launch from the end of 2013 through 2014, and at the date of writing (July 2014), 148 new gTLDs are available for general registrations, including 10 IDN new gTLDs. A further three IDN new gTLDs are at the sunrise period stage (ie the first stages of registration, reserved for trademark holders and others with pre-existing rights)¹⁰.

Figure 3 – IDN Introduction timeline





According to our research, as at July 2014, IDNs are offered at the second level under 50 ASCII TLDs.

37 IDN ccTLDs (for 26 countries and territories) have been added to the Internet root zone¹², of which 26 IDN ccTLDs have been launched, in addition to the original ASCII ccTLDs. This represents an increase of three since the same time last year. A further six are approaching the end of the approval process.

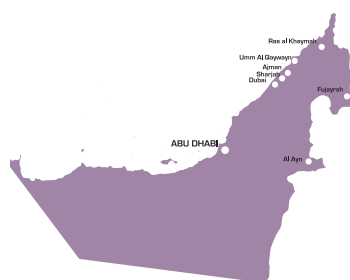
2013 also saw the launch of the first two IDN new gTLDs. Others followed in rapid succession and a total of 19 IDN new gTLDs have come to market.

APPENDIX 2

Country Case Studies

1 Arab States and Islamic Republic of Iran

1.1 United Arab Emirates¹



The .ae Domain Administration (.aeDA) was established in 2007 by the Telecommunication Regulatory Authority (TRA) as a department, regulatory body, and registry operator for .ae, the country code Top Level Domain (ccTLD) for United Arab Emirates (UAE). aeDA is responsible for setting and enforcement of all policies regarding the operation of the .ae ccTLD, and for overseeing the operation of the country's registry system².

When aeDA was first established, the registration of .ae domains was a manual process, which took at least 1 day to complete. The newly established registry implemented a liberalization strategy. Registration policies were relaxed (“no signature, no documentation, no local presence required”³), over 20 registrars have been accredited, and the registration process has now been automated so that domains can be in use within 2 minutes of placing an order. To support the liberalised registration system, a dispute resolution service has been implemented, operated by WIPO and closely modeled on the UDRP (the ICANN policy for handling disputes in .com and other generic Top Level Domains).

The registry has also conducted a sustained outreach programme, for example celebrating high profile users of .ae domain names, and using the Twitter hashtag #yes2ae. aeDA engages with its accredited registrars regularly in order to gain feedback on improving its service levels and customer satisfaction.

These changes have resulted in significant uplift in the number of ASCII domain names registered. At the end of 2013, the .ae TLD had over 112 000 registered domain names⁴. The uplift has not been felt in the .امارات IDN namespace, where there are only 2 200 registrations. The registry has been offering registrations in the IDN since April of 2010⁵.

The registry operator reports that the UAE experienced a typical IDN launch cycle along with a marketing campaign. User and registrar awareness of the IDN have been difficult to build. In addition, there was some confusion over how a native IDN was supposed to work with the more traditional ASCII names. It proved to be difficult to convince people that the IDN reflected an alternative way to reach Internet destinations.

Another barrier for the UAE IDN was poor support for Arabic in applications. Customers of the registry found that, if they typed a UAE IDN into a browser, the application converted it to Punycode. That translation of characters was viewed by some consumers as unexpected and possibly unwelcome behaviour by the browser.

However, the registry also found barriers in the ISP and hosting industry. If a customer had registered an IDN and wanted to use in for a website, the ISP or web hosting company had to know how to deploy the IDN. Early adopters found that the ISPs and infrastructure providers in the UAE were not yet prepared for the new development.

The registry reports that initiatives are underway to both improve the consumer view of the IDN as well as to help educate the Internet infrastructure industry on its use.

1.2 Qatar⁶



In 2010, the Qatar ccTLD registry changed operator to the Supreme Council of Information and Communication Technology⁷. The previous operator of the registry had run .qa as a closed registry, with restricted registration policies. The new registry changed this by introducing a new, first-come, first-served policy, with no limit on the number of domains registered by a single entity.

The IDN ending for Qatar, قطر, was launched at the same time as the relaunch of the .qa (ASCII) registry in 2010. The Qatar registry supported the launch of the IDN ccTLD with an aggressive marketing campaign. A three-month sunrise period was followed by a landrush.

At December 2013, there were 357 domains registered under قطر, a growth rate of 15% since December 2012. In relative terms this is a healthy annual growth rate, but in absolute numbers these are a small number of IDNs considering that the registry promotes them alongside the ASCII TLD.

In communication with the registry we found that they felt that they needed to break “the most common mindset of Registrants that fail to look beyond the traditional domain extensions.” They felt that this was due to “market perception” of the IDNs and that would change with increased awareness.

The registry also reports that homograph bundling is allowed. In fact, variants of the registered .qa IDN domains are first reserved upon request of the registrant. Up to five different variant domains can be registered for free.

Over the last three years, the Qatar registry has been building up its registrar base. It now has 15 registrars. The majority are international registrars, with only 3 based in the region. 12 of the registrars offer IDN registrations under قطر.

The Qatar registry has also been active in research and development, and in advocacy for the IDN. It has developed and launched a mobile app for registering both .qa and قطر domains.

1.3 Egypt



The National Telecom Regulatory Authority (NTRA) of Egypt is the operator of the مصر TLD. The TLDs مصر and .eg are operated by different organisations. The .eg domain has fairly low registration figures, and therefore limited visibility in the local market. Plans to raise awareness of the مصر domain through marketing campaigns have been put on hold due to political circumstances affecting Egypt. NTRA started registering IDNs in 2010. The IDN, xn--wgbh1c (مصر), was delegated in April 2010 – among the first four ccTLD IDNs to be approved and delegated through ICANN's IDN Fast Track process⁸.

44% of Egyptians are online. There are more than 12 million Facebook users, and use of social networks was cited as one of the causes of the Arab Spring uprising in Egypt. There is a vigorous Arabic online content, with Egypt as one of the top contributors. Mobile and fixed line broadband penetration are relatively healthy for the region.

Despite these promising factors, domain name registrations (both in ASCII and IDNs) exhibit the same low uptake seen across the Arab States. The aftermath of the Arab Spring in Egypt have meant that a scheduled land-rush phase for مصر was delayed until January of 2013.

The initial landrush was relatively successful. More than 3 000 domains were registered under the land rush strategy. However, a year later (February 2014), there were 3255 domains registered under مصر, a net growth of around 200 (8%) in the previous 12 months.

Despite relatively low registration figures, 52% of the registered مصر domain names have active name servers – indicating that the domain names are in active use. This is a higher percentage than many of the IDN ccTLDs in the study.

1.4 Saudi Arabia⁹



From 1995 to 2006 the Saudi ccTLD (.sa) was operated by the King Abdulaziz City for Science and Technology. In 2006, responsibility for the ccTLD was transitioned to SaudiNIC under the control of the Communication and Information Technology Commission (CITC).

All of SaudiNic’s services are offered and accessed electronically through its web portal. Most services are implemented automatically (without human interaction). The registry informs us that where certain services require human verification, such processes can be performed within 30 min (if the request is received within working business hours).

SaudiNIC does not operate a registry-registrar model. Nevertheless, numerous registrars offer Saudi domain names. However, retail prices are high (even though the registry charges no fee). SaudiNIC informs us that most registrants prefer to register their domain names directly with SaudiNIC, so the registrar price is unlikely to affect local registrants (which represents the majority of its customers).

In previous report, we have noted that the Saudi IDN domain, السعودية, had seen substantial growth rates in annual percentage terms. In December 2013, the number of registrations in .sa was 31 604 (an annual growth rate of 5% since December 2012), and the number of registrations in the Saudi IDN domain was 1 939 (an annual growth rate of 8% since December 2012).

SaudiNIC continues to be an important advocate for the Saudi IDN domain, السعودية, and has this year conducted extensive research on universal acceptance of Arabic domain names in popular applications and browsers¹⁰. SaudiNIC has also built a working model of an IDN email system which sends and receives Arabic email addresses¹¹. In response to the question “what single change would improve uptake of IDNs?” SaudiNIC told us “Support native IDN at the network layer (not as a hack on the application layer)”.

Since last year’s report, SaudiNIC has implemented a variant management scheme within the Saudi domains, and reports that, including variants, the number of IDNs under السعودية was 2 327 in May 2014. SaudiNIC is a member of ICANN’s Task Force on Arabic IDNs, which is working on Label Generation Rules for Arabic script Top Level Domains.

2 Asia and the Pacific

2.1 China¹⁸



The ccTLD registry for China, CNNIC, has been established and in continuous operation since 1997. At one stage, .cn was the largest ccTLD in the world, and had the same market share as .com in China. Registration numbers have fluctuated in recent years.

As well as managing the operation of .cn and 中国/中國, it conducts research and development and publishes an annual report on the development status of China's Internet¹⁹.

CNNIC was one of the first registries in the world to implement internationalised domain names (trial under .cn in 2000). There are still over 300 000 Chinese script domain names under .cn.

The IDN ccTLD 中国/中國 launched in 2010. CNNIC had offered a locally resolving version of these TLDs since 2000, but they did not work outside of China. In the first month of registration over 380 000 domains were ported over to the 中国/中國 domain. Over the subsequent two years, registrations decreased by 25%. However, through 2013, registrations have remained steady at 270 000.

The 中国/中國 TLD allows a combination of Chinese and ASCII characters. This is the only IDN ccTLD that the authors are aware of which allows ASCII registrations at the second level.

CNNIC operates a system for handling character variants, and as reported in last year's study, has been active in developing standards for IDNs in collaboration with others in the region. CNNIC has also been active in its advocacy for Chinese IDNs, with browser manufacturers, and in the field of internationalised email. During 2013, CNNIC has been working through the Internet Engineering Task Force to develop standards for POP and IMAP²⁰.

“政务”Government and Government Affairs and “公益” (Public Interest) new gTLDs
In April 2012, authorised by the Chinese Government, China Organizational Administration Center (CONAC), applied for two Chinese new gTLDs “政务” (government and

government affairs) and “.公益” (public interest) in ICANN's new gTLD programme. Both “.政务” and “.公益” TLDs launched in December, 2013.

Locally resolving TLDs

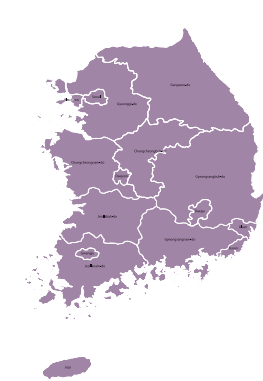
Before application of “.政务” and “.公益” gTLDs, CONAC had begun domestic operations of a testbed for tentative “.政务.cn”, “.政务”, “.公益.cn” and “.公益” TLDs since 2008. These worked within the networks of domestic operators only. The testbed has proved successful, and as of 30th May, 2014, the registration volume of “.政务.cn” and “.公益.cn” has reached 700 000.

Strict eligibility criteria apply for both the “.政务” and “.公益” TLDs.

IDN applications in browsers and IDN emails

CONAC has been promoting IDN applications in browsers in China and has cooperation with Baidu.Inc, the largest Chinese search engine, and has been researching IDN technology in Chinese-script email systems.


2.2 Republic of Korea²¹



In 2009, the Korean Internet and Security Agency was established by integrating three governmental agencies including National Internet Development Agency, which was responsible for IP addresses, DNS infrastructure and the .kr domain. The older KRNIC was incorporated into the new KISA to manage IP names and identifiers²². Among its duties, KISA is now responsible for both the .kr ccTLD and the 한국(.hankuk) IDN associated with the ccTLD.

The Korean IDN was launched with a sunrise period starting May 25, 2011 and a landrush period starting August 22, 2011. The general launch for the IDN started in October of 2011 and by the end of 2011 there were 210 000 한국 .hanguk domains registered. During 2012, the number of 한국 domain names reduced by over 50% to 91 000 due to lack of landrush renewals. 2013 has seen a continuation of the decline in the number of 한국 registrations: at the end of 2013 there were 60 000, a further reduction of 34%²³.

Much of the initial registration volume appears to have come from name speculation. At the end of 2011, only 29.45% of the names registered in the IDN resolved to a resource on the Internet. With the significant reduction in the number of domains registered in 2012, the number of IDNs that actually resolved rose to 55.17%. That compares to a resolution rate of 71.26% in the same time period for .kr domains. At the end of 2013, the resolution rate of 한국 .hanguk was 48.07% and the resolution rate of .kr domains was 70.37%, a slight reduction.



During 2013, KISA conducted research on the universal acceptance of 한국 domain names across mobile and desktop operating systems, applications including social networks, and browsers²⁴.

- All browsers (mobile and desktop) are successfully displaying the Korean IDNs, apart from Internet Explorer (lower than 6.0 (desktop) and Mobile Explorer 7.5)
- Applications popular in Republic of Korea (AL Tool, Naver and Google) accept Korean IDNs in the toolbar, both “테스트.한국” (test.hanguk), and “http://테스트.한국” (test.hanguk). Bing does not.
- During 2013, Naver and Daum (popular Korean applications) began to display 한국(hanguk) domain names in search results.
- Social Networks popular in Republic of Korea (Kakao, Naverline, and Nate On) have mixed results with supporting IDNs. Only Nate On supports the Korean IDNs fully.

KISA has also identified electronic mail as a critical requirement. KISA has constructed a test lab for examining the user environment for electronic mail combined with internationalised electronic mail addresses. KISA intends to implement a trial of Korean Internationalised email in 2014. KISA is also planning a cooperative project with Korean registrars in relation to internationalised email addresses, to take place in the second half of 2014.

As with many IDN registries, KISA is active in its advocacy. It is cooperating with the Korea Internet Corporations Association to raise user awareness of 한국 (hanguk) through the popular applications Naver, Daum, Kakao Talk and others. It continues to conduct research on the user experience of 한국 (hanguk) and on support for 한국 (hanguk) in popular applications.

KISA believes that collective effort is required to improve the universal acceptance of IDNs, involving the ICANN multistakeholder community and software developers; and that continual improvement in the support for IDNs and multilingualism across software, and programmes is necessary for the ICANN new gTLD programme to become a success.

2.3 Viet Nam²⁵



Until 2001, the incumbent Post and Telephone service for Viet Nam managed the ccTLD. From 2001, VNNIC, the Viet Nam Network Information Center took over management of .vn. VNNIC is affiliated with the Viet Nam Ministry of Information and Communications²⁶. As at October 2013, .vn had more than 260 000 registered names of which 66% had active websites²⁷. It is estimated that there were 214 000 total domain name registrations (across all top level domains) by Vietnamese organizations²⁸ suggesting that .vn has approximately 45% national market share. By December 2013, the number of “.vn” registrations had reached 266 028. The .vn domain is the largest in the ASEAN region and one of the largest ccTLDs in Asia region, behind Japan, and Republic of Korea²⁹.

The Vietnamese language is written in Latin script with diacritics, so, the .vn domain itself is meaningful for Vietnamese speakers.

The Vietnamese registry, VNNIC, conducted a trial of an IDN for .vn from 2004 to 2006. After the results of that trial were analysed, an official launch of the .vn IDN took place in March 2007. The official launch was limited to existing holders of ASCII .vn registrations.

In April of 2011, free and unlimited registration in the .vn IDN began. On the first day of the land rush 14 000 .vn IDNs were registered. During the first week there were 113 129 registrations under this policy. In the first four months there were 360,357 registrations. By October of 2013, the number of Vietnamese IDN registrations had reached 936 729, making .vn the largest IDN ccTLD registry in the world. It is estimated that IDN .vn domain names will reach 1 million registrations in June, 2014.

About 13% of registered IDNs combine with active services which are provided by VNNIC and its partner). While this is a low percentage, it is an increase of 5% on the previous year (9% in 2012). Analysis of active .vn IDNs indicates that 57% redirect to an existing web site (compared with 74% in 2012). A further 30% landed on a web hosting template.

VNNIC reports that in October 2013, four IDN .vn domain names received over 1 million queries, the most popular being “hàngloaviệtnam.vn” with 4.5 million queries³⁰.

3 Europe and North America

3.1 Russian Federation



The Russian Federation ccTLD operator, the Coordination Centre for TLD RU, is a not-for-profit company established in 2001. Following its foundation, the domain name registration system was substantially reorganised, and new accreditation processes introduced. In 2010, the Coordination Centre was delegated the Cyrillic IDN for the Russian Federation, .рф³².

In terms of volumes, the Russian Federation IDN TLD, .рф, remains the most successful of IDN experiences to date. When the landrush for .рф was opened in November 2010, 600 000 domains were registered in a single month. Each year, on the anniversary of launch, the landrush renewals phenomenon is felt³³, but each year the ripples diminish and the impact is lessened. Overall registrations continued at a healthy rate, with 811 000 .рф domains at December 2013, a net growth of 3.9% since the previous December.

Renewal rates of .рф domain names increased to 68% (from 61% during 2012). Renewal rates are seen in the industry as a long term measure of the value of a TLD to its users. While new registration rates may be distorted either through price promotions or speculation, renewal rates are tend to be linked to patterns of usage.

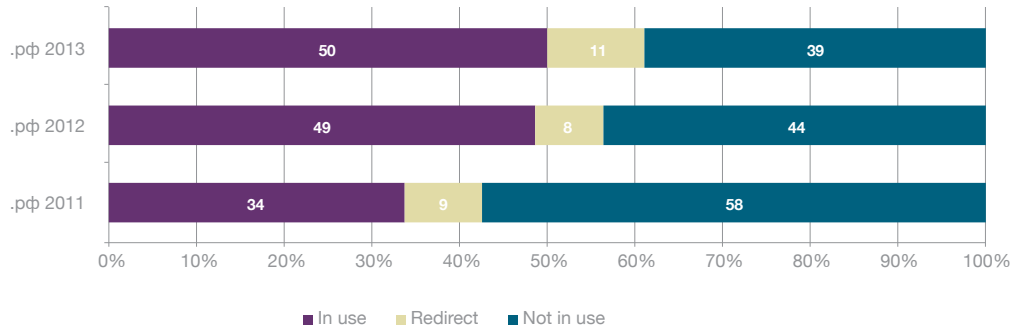
3.1.1 Usage

According to Statdom.ru³⁴, in December 2013 76% of .рф domains are delegated (ie capable of being used) up by 2% on the previous year. This still lags behind the delegation rate of over 90% in .ru.

Statdom also analyses the usage of .ru and .рф domains. In spite of the continuing challenges of using IDNs, overall usage of .рф domains has increased from 42% (December 2011) to 50% (2013). While usage rates are below those see in the ASCII .ru (69% in use), progress is encouraging (figure 4).



Figure 4 – Increased usage of .рф (2011-2013)



Usage rates of Russian IDNs are increasing year on year, an encouraging indication

The number of redirects is also increasing year on year, to 11% in 2013. This is above the level of redirects in the ASCII .ru (6% in 2013). IDN рф domains are increasingly being used in advertising, according to ccTLD.RU. When typed into a browser, the Cyrillic domain names redirect to an ASCII domain. In this way, it appears that the Russian market may be adapting to capture the marketing benefits of memorable, local script domain names, while using redirects to work around the currently unsatisfactory user experience of IDNs. For example, the photograph of the green and white van (figure 5) belongs to «ГрузовичкоФ» a cargo transportation company. The domain name the domain name грузовичкоф.рф advertised on the van redirects to <http://www.gruzovichkof.ru/>.³⁵

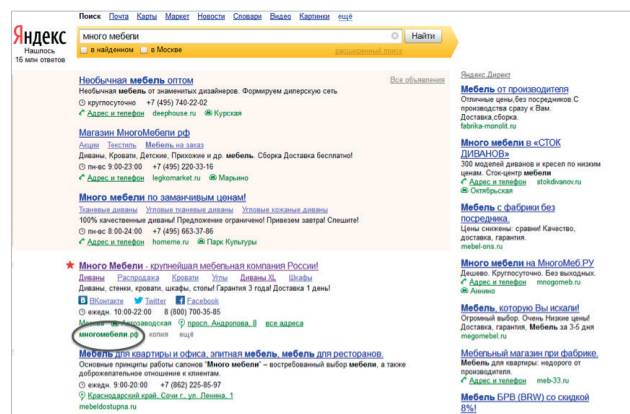
Figure 5



3.1.2 The user experience of Cyrillic domain names

Last year, the Russian registry reported that search engines were not prioritising Cyrillic URLs in their indexation. There has been an improvement in the last 12 months. Yandex, the most popular search provider in Russian Federation, now offers IDN search results. For example, in figure 6, the domain name shown on the red and white van is [многомебели.рф](http://mnogomebeli.ru). «Много мебели» means “a lot of furniture” in Russian, and is the name of a furniture manufacturing with a chain of retail stores. The domain name in this case does not redirect, and appears in the first page of search results through Yandex (figure 6).

Figure 6



The Russian registry has continued its tireless advocacy for IDNs during 2013:

- The registry has also written to Facebook to request that it updates its software to recognise Cyrillic websites. Facebook now does support links to IDN websites.
- It has had mixed success with Apple, whose iPhones carry a “.com” button. Following a request from the Russian registry, an IDN button was also included. However, this unfortunately disappeared in the next software release!

The Russian registry continues to report email as problematic for Cyrillic domains. The largest email portals are able to support sending email through their web pages. Unfortunately, web servers will not deliver the email. So, end to end the email sending is still unsatisfactory. However, it is reported the Mail.RU (a major webmail service provider) now supports IDN on the right of the @ sign (eg test@тест.рф) but does not support идн@идн, i.e. тест@тест.рф is not permitted.

ccTLD.RU identifies the biggest barrier to uptake of IDNs is lack of email functionality. If email were solved, the next issue is the keyboard. The @ sign is not a Russian character, requiring users to switch between Cyrillic and Latin keyboards when typing an email

Table 1 – Country and language factors

	China	Republic of Korea	Viet Nam	Islamic Republic of Iran	Egypt	Qatar	Saudi Arabia	United Arab Emirates	Russian Federation
UNESCO Region	Asia and the Pacific				Arab States				Europe and North America
GDP rank ³⁶	2	15	56	22	39	54	20	32	11
Literacy (secondary +) ³⁷	95%	96%	93.2%	99%	74%	96%	87%	89% ³⁸	99%
Cultural homogeneity ³⁹	High	High	High	High	High	Low	Moderate	Low	High
Official Language (Script)	Chinese (Han)	Korean (Hangul)	Vietnamese (Latin)	Persian (Arabic)	Arabic (Arabic)	Arabic (Arabic)	Arabic (Arabic)	Arabic (Arabic)	Russian (Cyrillic)
Linguistic homogeneity ⁴⁰	High-Moderate	High	High	High	High	Low ⁴¹	High	Low ⁴²	Moderate
Internet Exchange points ⁴³	4	4	3	0	2	0	1	1	16
Broadband Penetration ⁴⁴	13% (fixed)	37.5% 105.1%	4.4% 15%	4%	2.2% 21%	8.7% 61%	5.7% 40.4%	11% 21.7%	14.5% 183.5% ⁴⁵
Local language applications ⁴⁶	High ⁴⁷	High	--	--	High	Low	Low ⁴⁸	Low	--
Size of population ⁴⁹	1.35 bn	49 m	89 m	78 m	84 m	2 m	23.7 m	9.2 m	142 m
Online population ⁵⁰	46% ⁵¹	83%	35%	26%	44% ⁵²	88%	67%	71%	44.3%
IP address allocation ⁵³	High ⁵⁴	High	Low-Moderate	--	Low	High	Low-Moderate	Moderate-high	--
Overall country/language rating	Moderate-High	High	Low-moderate	Low-Moderate	Moderate-high	Low	Low-moderate	Low-moderate	High

Table 2 – ccTLD Factors⁵⁵

	China	Republic of Korea	Viet Nam	Islamic Republic of Iran	Egypt	Qatar	Saudi Arabia	United Arab Emirates	Russian Federation
UNESCO Region	Asia and the Pacific				Arab States				Europe and North America
Local registrars	High	High	Moderate	High ⁵⁶	Low-moderate	Low	None	Moderate	High
Policies	Eligibility criteria	Open	Open	Eligibility criteria ⁵⁷	Eligibility criteria	Open	Eligibility criteria	Open	Open
Pricing	Low	Low	Free (IDN)	-- Moderate (registrar)	---- High (registrar)	High	Free (registry) Moderate-High (registrar)	Moderate	Low
Brand strength	High ⁵⁸	High	Moderate	Moderate-High	Low-moderate	Low-Moderate	---	Moderate	High
Overall ccTLD factor rating	Moderate-high	High	Moderate	High	Low	Low-Moderate	Low	Moderate	High

APPENDIX 3

Methodology for language analysis

The source dataset is composed of 286 638 domain names from 40 top level domains that have Punycode (prefixed with xn--) components. The Unicode and Punycode values (if only one was available, the other was computed) were supplied by EURid (.eu only) or were extracted from the 39 relevant zone files publicly available in the first quarter of 2014.


Script: The second level domain name (in Unicode form) was examined. In it, the UTF-8 code point was determined for each character, the code point then compared against Unicode 6.3 Character Code Charts to determine the script, and then the employed scripts were aggregated for each domain.

Year: The created year was supplied by EURid for .eu domains and it has been kept in the comprehensive dataset. The creation date is not in the zone files and making whois lookups for all the other TLDs was not undertaken for privacy reasons. Therefore the year that is stored with the comprehensive dataset is that the year that the domains were extracted for this study from the respective zone files (December 2013 for .asia, .biz, info, .org; and March 2014 for those 35 TLDs starting with xn--).

IP Address: The IP Address and Reverse name was obtained by Linux utilities from the domain name (in Punycode form).

In use status: A domain must have an IP Address for it to display. Further, the web server will indicate by return code if the server believes it has successfully displayed the website page. Therefore, if the IP Address exists and the ultimate HTTP return code is between 200 and 299 (defined as Success) then the domain is considered "In use"; otherwise, it is considered "Not in use".

Redirect status: A domain must have an IP Address for it to display. Sometimes the page displayed is actually at another domain or at another IP address. If this has been done automatically by the website server, a special HTTP code is usually returned with the page. So if the IP Address exists and the immediate HTTP return code was between 300 and 399 (defined as Redirect) then the redirect status is considered "Redirected"; otherwise, it is considered "Not redirected". If there is no IP Address then the redirect status is considered "Cannot be determined". Please note: a domain could be redirected along a chain of IP addresses to an ultimate address that is either "In use" or "Not in use" (please see paragraph above).



Hosting country and ISP: IP2Location provides a database that contains the country and network provider that manages the network routing policy for IP Addresses. So if the IP Address exists for a domain in the dataset then that address was looked up in the database provided by IP2Location to determine the assigned country and managing network provider. If there is no IP Address then the country and ISP is considered "Cannot be determined".

Website language: The language names are those as maintained by IANA as indicated by the language code returned by Google Translate, the language code self-identified by the website, or as manually confirmed.

If the IP Address exists and content could be automatically downloaded, a sample (up to 200 characters if possible but no less than 30 characters) was extracted from the content from the <title>, <h1>, and <p> tags. If there were a sufficient number of characters in the sample then Google Translate was used to automatically identify the language. A random survey of the results showed that Google Translate was usually correct though occasionally wildly wrong whatever the sample's character count. We examined every language automatically "detected" by Google Translate and manually confirmed or corrected the outliers by opening the website in Chrome and employing Google Translate on the entire page. We considered a language to be an outlier if it was inappropriate for the TLD. For example, one outlier was an obscure African language that did not seem to match the TLD of .eu.

If there were less than 30 characters then the "lang" attribute was used, if it was available, to self-identify the language of a website. A random survey of the self-identified sites showed that they were either correct or they had used a country code instead of a language code. For example, a common mistake was to use se instead of sv for Swedish. We examined every self-identified language and manually confirmed the outliers. If there were too few characters in the sample and no "lang" attribute then the language is considered "Insufficient characters". If there is no IP Address then the language is considered "No content downloaded".

APPENDIX 4

Results of Usability Study 2014

	Website Ranking			Usability - Cyrillic Tests					
	eBizMBA	Alexa	Quantcast	CrAct	ConAct	Login	Prefs	Usability	Usability
Google	1	1	1	N	[N/A]	[N/A]	[N/A]	[N/A]	2
YouTube	2	2	3	N	[N/A]	[N/A]	[N/A]	[N/A]	2
Facebook	3	4	7	N	[N/A]	[N/A]	[N/A]	[N/A]	0
Yahoo	4	3	2	N	[N/A]	[N/A]	[N/A]	[N/A]	0
Amazon	5	6	18	N	[N/A]	[N/A]	[N/A]	[N/A]	0
Wikipedia	6	35	5	N	[N/A]	[N/A]	[N/A]	[N/A]	0
Twitter	7	10	8	N	[N/A]	[N/A]	[N/A]	[N/A]	0
Bing	8	27	11	N	[N/A]	[N/A]	[N/A]	[N/A]	0
eBay	9	7	6	N	[N/A]	[N/A]	[N/A]	[N/A]	0
msn	10	23	15	N	[N/A]	[N/A]	[N/A]	[N/A]	0
Pinterest			14	N	[N/A]	[N/A]	[N/A]	[N/A]	0
LinkedIn			16	N	[N/A]	[N/A]	[N/A]	[N/A]	0
PayPal			26	N	[N/A]	[N/A]	[N/A]	[N/A]	0

Guide to abbreviations

- CrAct The ability to create an account using an email address that uses an IDN as the fully qualified domain name (FQDN)
- ConAct The ability to confirm an account creation, usually done by email
- Login The ability to successfully log in using an IDN once an account has been created using the IDN as the FQDN
- Prefs Once logged into a web site, the ability to change preferences for the service
- Usability A subjective score (0: lowest; 10: highest) of how well the service supports IDNs and email addresses using IDNs

APPENDIX 5

Status of IDN ccTLDs (as at July 2014)

	ccTLD reference	Primary string	String in English	Script	Year of launch
United Arab Emirates	AE	xn--mg-baam7a8h امارات	Emarat	Arabic	2010
China	CN	xn--fiqs8S, 中国 xn--fiqz9S, 中國	China	Simplified Chinese, Traditional Chinese	2010 ¹
Algeria	DZ	xn--lgbbat-1ad8j الجزائر	Algeria/ Al Jazair	Arabic	2012 ²
Egypt	EG	xn--wgbh1c مصر	Egypt	Arabic	2010
Hong Kong SAR China	HK	xn--j6w193g 香港	Hong Kong	Han (simplified, traditional)	2011
India	IN	xn--h2brj9c भारत xn--mg-bbh1a71e بھارت xn--fpcrj9c3d ਝੰਡੇ xn--gecrj9c ଭାରତ xn--s9brj9c ভারত xn--45brj9c ভারত xn--xkc2d-13a5ee0h இந்தியா	Bharat / India	Devanagari (Hindi) Arabic Telugu Gujarati Gurmukhi (Punjabi) Bengali Tamil	Expected 2014

	ccTLD reference	Primary string	String in English	Script	Year of launch
Islamic Republic of Iran	IR	xn--mg-ba3a4f16a ایران	Iran	Arabic	Expected 2014
Jordan	JO	xn--mgbayh-7gpa الأردن	Al-Ordon	Arabic	2010
Kazakhstan	KZ	xn--80ao21a қаз	Kaz	Cyrillic	2012
Republic of Korea	KR	xn--3e0b707e 한국	Republic of Korea	Hangul	2011
Sri Lanka	LK	xn--fzc2c9e2c ලංකා xn--xkc2al3h-ye2a இலங்கை	Lanka llangai	Sinhala Tamil	2010
Morocco	MA	xn--mgb-c0a9azcg المغرب	Morocco / al-Maghrib	Arabic	Not yet launched
Malaysia	MY	xn--mgbx4cd-0ab مليسيا	Malaysia	Arabic	2012
Mongolia	MN	xn--l1acc МОН	Mon	Cyrillic	Expected 2014
Oman	OM	xn--mgb9awbf عمان	Oman	Arabic	2014
Palestine	PS	xn--ygb12am-mx فلسطين	Palestine	Arabic	2011
Qatar	QA	xn--wgb16a قطر	Qatar	Arabic	2011
Russian Federation	RU	xn--p1ai рф	rf	Cyrillic	2009
Republic of Serbia	RS	xn--90a3ac срб рб	srb	Cyrillic	2012
Saudi Arabia	SA	xn--mgbep4a5d4ar السعودية	Al Saudiah	Arabic	2010

	ccTLD reference	Primary string	String in English	Script	Year of launch
Singapore	SG	xn--yfro4i67o 新加坡 xn--clchc0e-a0b2g2a9gcd சிங்கப்பூர்	Singapore	Han Tamil	2011
Syrian Arab Republic	SY	xn--ogbpf8fl سورية	Syrian Arab Republic	Arabic	2011 (registrations not currently available)
Taiwan of China	TW	xn--kpry57d 台灣 xn--kprw13d 台灣	Taiwan, Province of China	Simplified Chinese Traditional Chinese	2010
Thailand	TH	xn--o3cw4h ไทย	Thai	Thai	2011
Tunisia	TN	xn--pgbs0dh تونس	Tunis	Arabic	2012 ³
Ukraine	UA	xn--j1amh укр	Ukr	Cyrillic	2013 ⁴

Pending delegation as at 14 July 2014⁵

	ccTLD reference	Primary string	String in English	Script
Bangladesh	BD	xn--54b7fta0cc বাংলা	Bangla	Bangla
Georgia	GE	xn--node გე	ge	Georgian (Mkhedruli)
Macedonia, Former Yugoslav Republic of	MK	xn--d1alf МКД	mkd	Cyrillic
Pakistan	PK	xn--mg-bai9azgp6j پاکستان	Pakistan	Arabic
Sudan	SD	xn--mgbp12fh سودان	Sudan	Arabic
Yemen	YE	xn--mgb2ddes اليمن	AlYemen	Arabic

Endnotes

World Report on IDN Deployment

- ¹ Source: UNESCO Atlas of Languages in Danger, <http://www.unesco.org/culture/languages-atlas/>, accessed 17 July 2014. More than 2 400 are classified as languages in danger
- ² From "English won't be the Internet's lingua franca", Oustinoff, M, 2011; *Net.Lang* 2012, p57
- ³ As measured on Lieberman's Diversity Index. See Yoshiki, M and Kodama, S, *Measuring Linguistic Diversity on the Net*, *Net.Lang* 2012, p132.
- ⁴ Source: W3Techs "Usage of content languages for websites" http://w3techs.com/technologies/overview/content_language/all, accessed 18 June 2014. The study reviews "only the top 10 million websites...We use the website popularity ranking provided by Alexa" See <http://w3techs.com/technologies>
- ⁵ Union Latina, *Fundredes* 2007, cited in Prado, D. "Language presence in the real world and Cyberspace", *Net.Lang* 2012.
- ⁶ Chinese is also becoming a lingua franca, but is not so popular as a language of online content.
- ⁷ Global Internet Report 2014, the Internet Society, http://www.internetsociety.org/sites/default/files/Global_Internet_Report_2014_0.pdf, accessed 7 July 2014.
- ⁸ Source: Ethnologue "Languages with at least 50 million first-language speakers" <http://www.ethnologue.com/statistics/size> accessed 18 June 2014.
- ⁹ See PRADO, 2012, *Net.Lang* p42
- ¹⁰ The W3Techs publish a useful analysis of the language of the world's 100 million most popular websites.
- ¹¹ Dubai School of Government, June 2013 http://www.dsg.ae/En/News/ASMR_5_Report_Final.pdf, accessed 18 June 2014.
- ¹² Sources: Ethnologue.com, W3Techs.com, Internetworldstats.com
- ¹³ Note that these are the languages supported in the application's environment, not the language of user generated content.
- ¹⁴ <https://about.twitter.com/company>, accessed 16 May 2014
- ¹⁵ <https://about.twitter.com/company/translation>, accessed 16 May 2014
- ¹⁶ http://translate.google.co.uk/about/intl/en_ALL, accessed 16 May 2014
- ¹⁷ <https://www.facebook.com/translations/FacebookLocales.xml>, accessed 16 May 2014; number excludes Leet speak, English (Pirate), English (upside down) and Latin.
- ¹⁸ <https://developers.facebook.com/docs/internationalization/>, accessed 16 May 2014.
- ¹⁹ <http://en.wikipedia.org/wiki/Wikipedia:Statistics>
- ²⁰ http://meta.wikimedia.org/wiki/List_of_Wikipedias, accessed 16 May 2014
- ²¹ [LECLERC, 2010] cited in *Net.Lang* 2012.
- ²² See "The Geographically Uneven Coverage of Wikipedia" Straumann, R., and Graham, M., Oxford Internet Institute, 2014 <http://geography.ox.ac.uk/?page=the-geographically-uneven-coverage-of-wikipedia> accessed 8 July 2014.
- ²³ Source: EURid UNESCO IDN World Report data
- ²⁴ Source: IEDR Annual Report 2012 (published 2013) <https://www.iedr.ie/wp-content/uploads/2014/04/IEDR-Annual-Report-Review-2012-B1.pdf>, accessed 8 July 2014
- ²⁵ Yandex has over 100 million users http://company.yandex.com/general_info/yandex_today.xml, accessed 16 May 2014.
- ²⁶ Nate On has 450 million users, according to KISA, presentation to APTLD, 11 May 2014 http://www.aptd.org/system/files/share/1/03_research_results_of_hanguk.pdf, accessed 16 May 2014
- ²⁷ Tencent QQ has 798.2 million users as of December 2012 <http://www.tencent.com/en-us/at/abouttencent.shtml>, accessed 16 May 2014.
- ²⁸ See Global Internet Report, the Internet Society http://www.internetsociety.org/sites/default/files/Global_Internet_Report_2014_0.pdf, and <http://w3techs.com/technologies> for an explanation of the methodology.
- ²⁹ There are a few exceptions eg Tokelau, China, Russia, Brazil. The .tk domain (Tokelau) has been pursuing a "giveaway" policy for the past few years, resulting in high volumes of registrations. Despite the success of the ccTLDs of China, Russia and Brazil, some European ccTLDs enjoy a far higher penetration rate in their local populations (eg Netherlands).
- ³⁰ Statdom.ru. <http://statdom.ru/tid/%D1%80%D1%84/report/sitesusage/#26>, accessed 30 June 2014. For usage rate of pdf as at December 2013, we have discounted "No IP address, 12%" and "Not delegated, 22%"
- ³¹ *The Relationship between local content, internet development and access prices*. Internet Society, OECD, and UNESCO, 2011.
- ³² Source: Ethnologue
- ³³ EURid quarterly report for Q4 2013, http://www.eurid.eu/files/publ/Q4_2013.pdf
- ³⁴ <http://blogs.informatandm.com/17551/press-release-the-total-number-of-ite-subscriptions-will-reach-1-3-billion-by-the-end-of-2018/>
- ³⁵ http://www.gsacom.com/downloads/pdf/Global_Mobile_Broadband_Market_Update_160514_small.php4

- ³⁶ For instance, using http://www.google.com/advanced_search works as expected, but Microsoft's Bing produces no results for the same search.
- ³⁷ Source: Presentation to APTLD meeting, Oman, May 2014, http://www.aptdl.org/system/files/share/1/03_research_results_of_hanguk.pdf, accessed 17 June 2014.
- ³⁸ Which appears to be partly the result of using tools that allow for text input fields (for instance, entering the text for a Tweet) to be encoded in UTF-8.
- ³⁹ SSL is Secure Sockets Layer and TLS is Transport Layer Security. These are technologies that make it possible to encrypt communications between sender and receiver – making the communications channel more secure.
- ⁴⁰ For instance, see <http://tools.ietf.org/html/rfc6763> for a description of a service discovery protocol using the DNS that requires the service names to be in ASCII.
- ⁴¹ An overview is available at: <http://tools.ietf.org/html/rfc6530>
- ⁴² "Raseel - Arabic Emails," Presented at the APTLD Meeting, Oman, May 2014, <http://www.citc.gov.sa>
- ⁴³ <http://datatracker.ietf.org/doc/rfc6532/>
- ⁴⁴ <http://tools.ietf.org/html/rfc6530>
- ⁴⁵ Source: <http://emailclientmarketshare.com>, accessed 12 July 2014
- ⁴⁶ Are you able to configure a POP, IMAP and SMTP account using International Email Addresses?
- ⁴⁷ Are you able to send an email to your nearest email server using an Internationalised Email Address in the "To:" field?
- ⁴⁸ "Raseel - Arabic Emails," Presented at the APTLD Meeting, Oman, May 2014, <http://www.citc.gov.sa>
- ⁴⁹ As noted in the CNNIC Annual Report, <http://www.cnnic.net.cn/gwym/zkzw/cnnicnbg/201401/U020140314621129945634.pdf>
- ⁵⁰ CNNIC Press Release, undated, "Postfix Will Support IETF IEA Protocol this year"
- ⁵¹ For a description of the problem posed by homograph attacks, see <https://www.icann.org/news/announcement-2005-02-23-en>; accessed June 2014
- ⁵² Processing an IDN URL as a whole means examining the entire IDN string as a single unit. Processing an IDN URL label-by-label means looking at each label (i.e. the strings separated by the dot ".") individually and making a display decision for each separate component of the string.
- ⁵³ Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1–55.
- ⁵⁴ The Relationship between local content, Internet development and access prices", *Internet Society, OECD and UNESCO*, 2011.
- ⁵⁵ *ibid*
- ⁵⁶ UNESCO World Report on Cultural Diversity, 2009
- ⁵⁷ World Values Survey database, 2008, referenced in UNESCO World Report on Cultural Diversity 2009, table 6
- ⁵⁸ Source: ITU Key 2005-2014 ICT data http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2014/ITU_Key_2005-2014_ICT_data.xls accessed 20 August 2014
- ⁵⁹ Dubai School of Governance, Arab Social Media Report, 2013 http://www.dsg.ae/En/News/ASMR_5_Report_Final.pdf
- ⁶⁰ *ibid*
- ⁶¹ For this comparison, we included Islamic Republic of Iran in the Arab States, because of its use of Arabic script.
- ⁶² Arabic language is the 7th most popular language used on the Internet (2011), with 2 500% growth between 2000-2011 (source: registry presentation to Middle East DNS Forum, February 2014).
- ⁶³ Source: Registry presentation to Middle East DNS Forum, February 2014.
- ⁶⁴ See إدارة أسماء نطاق الإنترنت (aeDA.) 100 ألف نطاق (ae.) تحقق رقم 100 <http://aeda.ae/ar/news-ar.php?id=107> accessed 8 July 2014
- ⁶⁵ See <http://www.bqdoha.com/2013/12/population-qatar>, accessed 16 June 2014
- ⁶⁶ Thanks to Sarmad Hussain and Fahd Batayneh for the information in this section.
- ⁶⁷ See <https://community.icann.org/display/MES/TF-AIDN+Work+Space>
- ⁶⁸ The source for this section is ntldstats.com, accessed 8 May 2014.
- ⁶⁹ Our analysis excluded brands, geographic indications, transliterations of .com, .net and .org, and community applications. We counted multiple applications (eg. .app, .shop) for the same string on the basis that they indicated popularity of certain terms
- ⁷⁰ Source Ethnologue: <http://www.ethnologue.com/statistics/size>, and ICANN gTLD applications.
- ⁷¹ According to domainincite.com, one applicant estimated year 1 revenues of \$28.4 million. <http://domainincite.com/13929-new-gtld-revenue-projections-revealed-in-leaked-famous-four-presentation>.
- ⁷² According to public portions of its applications, Donuts, which has applied for hundreds of new gTLDs, raised over \$100 million in venture capital. <http://www.donuts.co/tlds/>, and <https://gtldresult.icann.org/application-result/applicationstatus/applicationdetails/736>
- ⁷³ Source: Presentation of ICTQatar, ICANN ME DNS Forum, <https://icann.adobeconnect.com/p6tcakxpr9/?launcher=false&fcsContent=true&pbMode=normal> accessed 5 June 2014
- ⁷⁴ DotShabaka CEO, Yasmin Omer, quoted in "شبكة. Becomes the world's first Arabic borderless Top-Level Domain, 14 July 2013, <http://www.dotshabaka.com/media-detail-en.php?2013-07-14-00-00-00-1> accessed 7 July 2014.
- ⁷⁵ Source ntldstats.com <http://ntldstats.com/tld/xn--ngbc5azd>, accessed 7 July 2014.
- ⁷⁶ See <http://domainincite.com/16016-google-giving-away-5000-free-new-gtld-domains>, accessed 8 May 2014
- ⁷⁷ <http://xn--p8j9a0d9c9a.xn--q9jyb4c/index.html>, accessed 8 May 2014.

- ⁷⁸ Source: <http://ntldstats.com/tld/xn--q9jyb4c>, accessed 8 May 2014.
- ⁷⁹ See <http://www.interlink.or.jp/>
- ⁸⁰ Source: <http://ntldstats.com/tld/xn--fiq228c5hs> and <http://ntldstats.com/tld/xn--3ds443g> and, accessed 7 July 2014.
- ⁸¹ Source: <http://ntldstats.com/registrar/1387-1API-GmbH>, accessed 7 July 2014.
- ⁸² The data sample quoted in previous reports was 228 million (90%) for 2012, 203 million for 2011, and 163.7 million for 2010. We have backfilled previous years' data when new information became available.
- ⁸³ Note: several ccTLDs have implemented more than one IDN TLD. There are various reasons for this, for example handling of character variants (China, Taiwan of China) or to support local languages which use different scripts (Sri Lanka, Hong Kong SAR China, Singapore).
- ⁸⁴ Note that the data sample for 2013 is larger than in previous years, as more ccTLD collaborate in this project, so this chart shows the totals in our data sample in the two years, and not necessarily the same as the growth rate.
- ⁸⁵ There were 1.9 million in 2009, and by 2013 this had grown to 6 million.

Appendix 1

- ¹ For an explanation of ASCII, see <http://en.wikipedia.org/wiki/ASCII>, accessed 21 August 2012.
- ² See *Internationalization of Domain Names: A history of technology development*, Klensin, J and Fältström, P.
- ³ See <http://www.icann.org/en/topics/idn/idn-activities-seoul-28oct09-en.pdf>.
- ⁴ IDN variants have been a focus for working groups within ICANN recently. See <http://www.icann.org/en/news/public-comment/idn-variant-tld-revised-program-plan-04may12-en.htm>, accessed 16 May 2012.
- ⁵ See address of Janis Karklins, Assistant Director General, Communications and Information Sector, UNESCO, at the Opening Ceremony of the IGF Vilnius 2010.
- ⁶ See "The relationship between local content, Internet development and access prices", Internet Society, OECD and UNESCO, 2011.
- ⁷ Lipsicas, B. and Shikmoni, D., "Internationalized Domain Names: the long and winding road", *CENTR Domain Wire*, Issue 1, 2007.
- ⁸ See Internet Governance Forum 2006, Diversity main session, <http://www.intgovforum.org/cms/IGF-SummingUp-0111106.txt>, accessed 23 May 2012.
- ⁹ See <http://www.icann.org/topics/idn/idn-activities-seoul-28oct09-en.pdf>.
- ¹⁰ Source: www.ntldstats.com, accessed 7 July 2014.
- ¹¹ <http://www.icann.org/en/resources/idn/fast-track>
- ¹² <http://www.icann.org/en/resources/idn/fast-track/string-evaluation-completion>, accessed 23 April 2012, and <https://charts.icann.org/public/index-idn.html>, accessed 17 June 2012.

Appendix 2

- ¹ Source: Nations Online http://www.nationsonline.org/maps/united_arab_emirates_map.jpg, accessed 28 July 2013.
- ² Source: .aeda, About Us, <http://www.aeda.ae/eng/aboutus.php>, accessed 29 July 2013
- ³ Source: aeDA promotional video (shown at ICANN Middle East DNS Forum <https://icann.adobeconnect.com/p354gr0pbh4/>, accessed 5 June 2014)
- ⁴ Source: aeDA presentation to ICANN Middle East DNS Forum <https://icann.adobeconnect.com/p354gr0pbh4/>, accessed 5 June 2014
- ⁵ Source: .emarat (إمارات) IDN confirmed for April 2010, <http://www.domaining.ae/2009/11/amarat-%D8%A7%D9%85%D8%A7%D8%B1%D8%A7%D8%AA-confirmed-for-april-2010/>, accessed 1 August 2013
- ⁶ Source: Nations Online <http://www.nationsonline.org/maps/qatar-map.jpg>, accessed 28 July 2013.
- ⁷ Source: Qatar Regulatory Authority Annual Report 2012, http://www.ictqatar.qa/sites/default/files/documents/RA_Annual_Report_2011_EN_1.pdf, accessed 25 July 2013.
- ⁸ Source: IDN Fast Track Process, <http://www.icann.org/en/resources/idn/fast-track>, accessed: 24 July 2013.
- ⁹ Source: Nations Online <http://www.nationsonline.org/maps/Saudi-Arabia-Map.jpg>, accessed 28 July 2013.
- ¹⁰ See presentation "Assessment of IDNA user experience", by Mohammed Alhamed to APTLD meeting, May 2014, <http://nic.sa/docs/ADN/IDNAAssessment-v3-raed.pdf>, accessed 10 June 2014
- ¹¹ See presentation "Arabic emails" by Mohammed Alhamed to APTLD Meeting, May 2014, <http://nic.sa/docs/ADN/raseel.pdf>
- ¹² <http://www.nationsonline.org/oneworld/index.html>
- ¹³ Source: https://www.nic.ir/Terms_and_Conditions_iran, clause 6, accessed 5 July 2013.
- ¹⁴ Source: https://www.nic.ir/Dispute_Resolution_Policy_ir, accessed 5 July 2013.

- ¹⁵ Source: https://www.nic.ir/Dispute_Resolution_Providers_ir, accessed 5 July 2013.
- ¹⁶ Source: direct communication with IRNIC. For more information on homoglyph bundling under .ir, see https://www.nic.ir/Allowable_Characters_dot-iran#Table_3, accessed 5 July 2013.
- ¹⁷ Source: <http://www.nic.ir/Statistics>, accessed 18 June 2014.
- ¹⁸ Source: <http://www.nationsonline.org/oneworld/>, accessed 28 August 2013
- ¹⁹ eg Statistical Report on Internet Development in China (January 2014) <http://www1.cnnic.cn/IDR/ReportDownloads/201404/U020140417607531610855.pdf>, accessed 4 June 2014.
- ²⁰ http://www1.cnnic.cn/AU/MediaC/rdxw/hotnews/201304/t20130416_39287.htm
- ²¹ Source: Nations Online <http://www.nationsonline.org/maps/korea-south-admin-map.jpg>, accessed 1 August 2013.
- ²² Source: Chronology, <http://www.kisa.or.kr/eng/aboutkisa/chronology.jsp>, accessed 2 August 2013
- ²³ Source: Domain PR and Marketing Results of 2012, independent paper obtained from author
- ²⁴ Source: Presentation to APTLD meeting, Oman, May 2014, http://www.aptd.org/system/files/share/1/03_research_results_of_hanguk.pdf, accessed 17 June 2014.
- ²⁵ Source: Nations Online <http://www.nationsonline.org/maps/vietnam-political-map.jpg>, accessed 1 August 2013.
- ²⁶ Source: VNNIC, <http://www.vnnic.vn/en/>, accessed 4 August 2013
- ²⁷ Source VNNIC: Report on Vietnam Internet Resources 2013, <http://www.vnnic.vn/sites/default/files/tailieu/whitebookfinal%28versionE%29.pdf>, accessed 17 June 2014
- ²⁸ Ibid. Note that the number in last year's report was incorrectly stated as 850 000
- ²⁹ Ibid
- ³⁰ Ibid, accessed 4 August 2013
- ³¹ Source: extracted from http://www.nationsonline.org/oneworld/asia_map.htm, accessed 28 August 2013.
- ³² Source: <http://cctld.ru/en/about/> Coordination Center for TLD RU website, accessed 24 May 2012.
- ³³ Source <http://statdom.ru/tld/%D1%80%D1%84/report/domainscount/#8:by=month>, accessed 17 June 2014.
- ³⁴ Source: <http://statdom.ru/tld/%D1%80%D1%84/report/domainscount/#8:date=20130101-20131201&by=month> accessed 17 June 2014.
- ³⁵ Thanks to Irina Danelia of ccTLDRU for the photographs
- ³⁶ World Bank <http://data.worldbank.org/data-catalog/GDP-ranking-table> (as at 2012, accessed 4 June 2014)
- ³⁷ World Bank Education Statistics <http://databank.worldbank.org/data/views/reports/tableview.aspx?isshared=true> (queried 4 June 2014),
- ³⁸ Source <http://www.uaecd.org/education-introduction>, accessed 4 September 2013
- ³⁹ This measure is drawn from a number of indicators in UNESCO World Report on Cultural Diversity 2009 <http://unesdoc.unesco.org/images/0018/001852/185202e.pdf>, including migrant stock, national pride, outbound and inbound student flows, cultural home production imports and exports, tourism inbound and outbound flows
- ⁴⁰ Ethnologue <http://www.ethnologue.com> (provides information by country), queried 4 June 2014. Where the number of individual languages is greater than 20, the measure derived from the percentage of individual languages classified as "institutional". For example, in China, 298 languages, of which 5% are institutional. Where a country has a high percentage of immigrant languages, this is recognised in the evaluation.
- ⁴¹ According to UNESCWA, international migrants make up 74% of the Qatari population. See <http://www.escwa.un.org/popin/members/qatar.pdf>, accessed 4 June 2014
- ⁴² According to UNESCWA, international migrants make up 84% of the UAE population (2013) See <http://www.escwa.un.org/popin/members/uae.pdf>, accessed 4 June 2014
- ⁴³ Packet Clearing House <https://prefix.pch.net/applications/ixpdir/summary/>, plus interviews with ccTLD operators in country
- ⁴⁴ ITU time series by country 2000-2013, and aggregate data 2006-2013 <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>, and The Broadband Commission, The State of Broadband 2012: Achieving Digital Inclusion for All, <http://www.broadbandcommission.org/Documents/bb-annualreport2012.pdf>
- ⁴⁵ 2012 figure for CIS region, Source: ITU Key Aggregate Data 2005-2013 <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- ⁴⁶ Source: Interviews with ccTLD operators in country: ISOC, OECD, UNESCO study on Local Infrastructure and Local Content 2011
- ⁴⁷ CNNIC Statistical Report, 2013 <http://www1.cnnic.cn/IDR/ReportDownloads/201302/P020130221391269963814.pdf>
- ⁴⁸ The Broadband Commission, The State of Broadband 2012: Achieving Digital Inclusion for All, <http://www.broadbandcommission.org/Documents/bb-annualreport2012.pdf>,
- ⁴⁹ World Bank, country data (last updated 2012), accessed 4 June 2014 <http://data.worldbank.org/country>
- ⁵⁰ ITU time series by country 200-2012, and aggregate data 2006-2013 <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- ⁵¹ CNNIC, 33rd Statistical Report on Internet Development in China (January 2014) <http://www1.cnnic.cn/IDR/ReportDownloads/201404/U020140417607531610855.pdf>, accessed 4 June 2014
- ⁵² Source: Egyptian Ministry of ICT, presentation to ICANN Middle East DNS Forum, February 2014
- ⁵³ IPv6 Test, <http://ipv6-test.com/stats/country/SA>
- ⁵⁴ CNNIC, 33rd Statistical report, p32-33

⁵⁵ Source: Interviews with ccTLD operators; ccTLD websites; registrars' websites.

⁵⁶ IRNIC has 45 locally accredited resellers (registrars) and 9 international resellers (registrars). See <https://www.nic.ir/Statistics>, accessed 10 June 2014.

⁵⁷ Proof of identity is required.

⁵⁸ According to CNNIC, 33rd Statistical Report on Internet Development in China (January 2014) <http://www1.cnnic.cn/IDR/ReportDownloads/201404/U020140417607531610855.pdf>, accessed 4 June 2014, there are a total of 18 million domain names in China, of which nearly 60% are .cn domains. The .cn namespace saw 44% growth in the year 2012

Appendix 5

¹ http://www.cnnic.cn/jczyfw/zwym/zwymtjxx/201206/t20120612_26518.htm, accessed 13 July 2013

² <http://xn--ggbdmbaav3cjl1c9heugfv.xn--lgbbat1ad8j/>, accessed 13 July 2013

³ <http://www.registre.tn/en/index.php?rub=262&sru=360> accessed 7 July 2014

⁴ http://www.ukrinform.ua/eng/news/ukraine_among_50_countries_with_national_domain_names_308223

⁵ <http://www.icann.org/en/resources/ids/fast-track/string-evaluation-completion>



Glossary of terms

- **ASCII**

The American Standard Code for Information Interchange, representing text in computers, communications equipment and other devices. In the context of the domain name system ASCII means the letters "a-z" inclusive, the numerals "0-9" inclusive and the hyphen "-". Until the year 2000, no other characters were allowed in domain names, and in 2009, the first IDN ccTLDs were introduced
- **ccTLD**

Country code Top Level Domain, which represents a country or territory found in the ISO 3166 list, for example .eu (European Union), .de (Germany), .uk (United Kingdom), .fr (France).
- **CENTR**

The European country code Top Level Domain organisation, a not-for-profit organisation which supports the interests of ccTLD managers.
www.centri.org
- **EURid**

The European Registry of Internet Domain Names, EURid, manages the .eu top level domain under contract to the European Commission. The .eu TLD was launched for general registration in 2006, and has over 3.6 million domain names.
- **gTLD**

Generic Top Level Domain, which does not represent a particular country or territory. Examples include .com, .net, .org, .info, and .biz.
- **Hybrid IDN, hybrid domain**

An internationalised domain name in which the constituent elements are in different scripts. Examples of hybrid IDNs are shown in Appendix 1, figure 2.
- **ICANN**

The Internet Corporation for Assigned Names and Numbers. A non-profit company responsible for management of the domain name root operation (the IANA), policy coordination for generic Top Level Domains (gTLDs), and for Internet numbering. In 2012, ICANN launched a process to create an unlimited number of new gTLDs, over 1 900 applications were received. ICANN's policy development is guided by a number of support organisations and advisory committees representing various stakeholder groups including governments, the domain name industry, business, ccTLD registries, and civil society.
www.icann.org
- **IDN**

Internationalised Domain Name. A domain name written in non-Latin scripts such as Chinese, Arabic, Hangul, or Cyrillic. For an explanation of IDNs, see Appendix 1.

- **IDN ccTLD**

A country code domain written in non-Latin scripts. Examples include 한국 (Republic of Korea), قطر (Qatar), 中国 (China), .рф (Russian Federation).
- **IDN ccTLD Fast Track**

A process developed within ICANN by the ccTLD registries to implement IDN ccTLDs. The first IDN ccTLDs were approved by ICANN in 2009. The IDN ccTLD Fast Track process continues, and to date 31 IDN ccTLDs have been approved by ICANN, of which 19 have launched for public registrations. The remainder are preparing to launch.
- **IETF**

Internet Engineering Task Force. Develops Internet standards. Its members are volunteers from the international technical community, and it is open to any interested individual. IETF standards are published as Requests for Comment (RFC). www.ietf.org
- **ISOC**

The Internet Society. Formed in 1992, it promotes the open development, evolution and use of the Internet for all. www.isoc.org
- **ISP**

Internet Service Provider. An organisation that provides access to the Internet, and a variety of related services including web hosting, or email services.
- **IXP**

Internet Exchange Point. Internet Service Providers (ISPs) can exchange Internet traffic between their networks, thereby reducing costs and increasing speed in resolving Internet queries (eg web pages).
- **Landrush**

When a new TLD is first launched, there is a period of time when trademark holders and others who have rights in particular names or brands have the opportunity to pre-register domain names (Sunrise Period). Following the Sunrise period, the registry opens to general registrations – this is called the landrush.
- **OECD**

Office for Economic Co-operation and Development. www.oecd.org
- **Punycode**

The syntax by which a string of Unicode characters is transliterated uniquely and reversibly into the ASCII character set used by the Domain Name System. Punycode is the underlying technology which makes IDNs possible. See section 1 for further explanation.



- **Register**

The domain name database managed by a registry.

- **Registrant**

A domain name registrant is the person or organisation in whose name or on whose behalf a domain name is registered. For example, the British Broadcasting Corporation (BBC) is the registrant of the domain name `bbc.co.uk`.

- **Registrar**

A domain name registrar. An organisation that is allowed to register domain names in one or more TLDs on behalf of its customers. To register in gTLDs, registrars must be accredited by ICANN; some ccTLDs operate their own systems of registrar accreditation. Examples of well-known registrars are Go Daddy, Inc, Tucows, and 101Domains.com.

- **Registry**

A domain name registry is a Top Level Domain provider, for example EURid is the registry for `.eu`, Verisign for `.com`.

- **Second level domain**

Domain names have a hierarchical structure, starting (in left to right scripts) to the right of the dot, with the Top Level Domain. Most domain names are registered at the second level, eg under `.eu`, or `.com`. In a domain name `example.com`, "example" is a second level domain. Some domains, eg `.uk` and `.jp` only register domain names at the third level, eg under `.co.uk`, or `.co.jp`.

- **TLD**

Top Level Domain. The domain name system is hierarchical, and is organised into various Top Level Domains (TLDs), eg `.com`, `.eu`, `.pф` under which domain names can be registered.

- **UNESCO**

United Nations Educational, Scientific, and Cultural Organisation, whose mission is building peace in the minds of men and women. UNESCO is organised into four sectors, including Communication and Information Sector whose mission is Building inclusive knowledge societies through information and communication.
www.unesco.org

- **Unicode**

A technical standard used for consistent encoding of text from ASCII into other scripts.

- **WSIS**

The World Summit on the Information Society, a UN process which took place in two phases 2003 and 2005, and resulted in the Geneva Declaration of Principles, Geneva Plan of Action, the Tunis Commitment and the Tunis Agenda. A number of UN organisations, including UNESCO, have been tasked with fulfilling action lines resulting from the WSIS.

Acknowledgements

Main author, Emily Taylor, with close collaboration of Giovanni Seppia of EURid, and Irmgarda Kasinskaite of UNESCO, Patrick Kane and Dennis Tan Tanaka of Verisign, Mark McFadden (author of Universal Acceptance chapter), Oxford Information Labs Ltd (IDN database and world maps), Giovanna Marotta and Margherita Donati of University of Pisa (authors, Reading, visual perception and language competence chapter).

The authors thank all those from the domain name industry and technical community who have generously contributed information and comments to this report. Special thanks to the regional ccTLD organisations and all their members: Don Hollander of APTLD, Carolina Aguerre of LACTLD, Peter van Roste of CENTR. Thanks to Mohammed Alhamed and Raed Al-Fayez of SaudiNIC, Hongbin Zhu of CNNIC, Limei Liu of CONAC, Minjung Park, Sin Yeong Park of KISA, Nguyen Thu Thuy of VNNIC, Irina Danelia of ccTLD.RU, Christine Arida and Manal Ismail of NTRA Egypt, Alireza Saleh of IRNIC, Mohammad al Zarooni, Tariq Mirza Ali, Hamad Al Zaabi of AEDA, Mohammed el Bashir of ICTQatar, Jonathan Shea of HKIRC, Baher Esmat, Fahd Batayneh and Sarmad Hussain of ICANN, Yasmin Omer of DotShabaka.

Disclaimer

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of EURid and UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

In addition, the co-publishers (EURid and UNESCO) have made every effort to ensure that the information contained in this publication is correct and current at the time of publication but take no responsibility for any inaccuracy.

The authors are responsible for the choice and the presentation of the facts contained in this publication and for the opinions expressed therein, which are not necessarily those of EURid or UNESCO and do not commit those organisations.

EURid vzw/asbl (headquarters)
Woluwelaan 150, 1831 Diegem, Belgium
T: +32(0)2 401 27 50
F: +32(0)2 401 27 51