POLICY PAPER

Bolstering Internationalised Domain Name Uptake in India

UJJWAL KRISHNA

May 2021

Table of Contents

List of Tables

List of Figures

Acknowledgements

We are grateful to the Ministry of Electronics and Information Technology (MEITY), Government of India, and the National Internet Exchange of India (NIXI) for the opportunity to conduct policy-relevant research under the ICANN Research and Multistakeholder Engagement Assistance Programme. We would like to thank their teams for their support.

Alexandra Tristant provided excellent research assistance and her contribution is gratefully acknowledged.

Finally, we are grateful to our team at ICRIER for their generous support and helpful ideas. All errors remain our own.

Bolstering Internationalised Domain Name Uptake in India

Ujjwal Krishna

1. Introduction

The year 1991, which saw the initiation of India's economic liberalisation, was also the year the World Wide Web was made available to the public. It was not until August 15, 1995 that the Internet was launched in India by the then state-owned Videsh Sanchar Nigam Limited (VSNL), a time when around 73 percent of Indians lived in rural areas (World Bank 2018). Despite rapid rates of economic growth since then, spurred on by the growth impacts of a newly liberalised economy, urbanisation has not proceeded on the expected lines with nearly 66 percent of India's population continuing to live in rural areas (World Bank 2018)[1], making it the country's largest population segment.

There are over 451 million active Internet users aged 5 and above (Nielsen-IAMAI 2019), and internet penetration stands at an unacceptably low 36 percent, meaning that roughly more than one in three individuals aged 12 and above are accessing the Internet in India. Internet penetration in urban areas is 51 percent vis-à-vis a meagre 27 percent in rural areas (ibid). This divide, with urban Internet penetration being around twice that of rural areas, is somewhat mirrored in the severe economic inequality that has come to characterise India's post-liberalisation economy, where almost 63 percent of the top 10 percent rich Indians reside in urban India, while 85 percent of the bottom-segment households reside in rural India (Oxfam 2019).

With a growth model that is increasingly exacerbating economic inequalities, bringing rural India online in a meaningful way is crucial to ensure sustainable and equitable development. The attendant benefits of being online and participating in the digital economy have further improved urban India's quality of life and incomes. The chasm between urban and rural prosperity will only increase at a more rapid pace if relevant strategies to on-board rural India to the digital world are not carefully conceived of, and effectively implemented.

At the same time, India is also poised to become the youngest country in the world with an average age of 29 years by 2020 (UN Habitat, 2013)[2], and will account for 28 percent of the world's workforce (EY and FICCI, 2013)[3]. This is popularly referred to as the 'demographic dividend', which is 'the economic growth potential that can result from shifts in a population's age structure, mainly when the share of the working-age population (15 to 64) is larger than the non-working-age share of the population (14 and younger, and 65 and older)' (UNFPA 2019)[4]. India's population pyramid is expected to 'bulge' in the aforementioned working-age

1 https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=IN

2 UN Habitat. (2013). *State of the Urban Youth 2012-2013: Youth in the Prosperity of Cities.* Nairobi: United Nations Human Settlements Programme

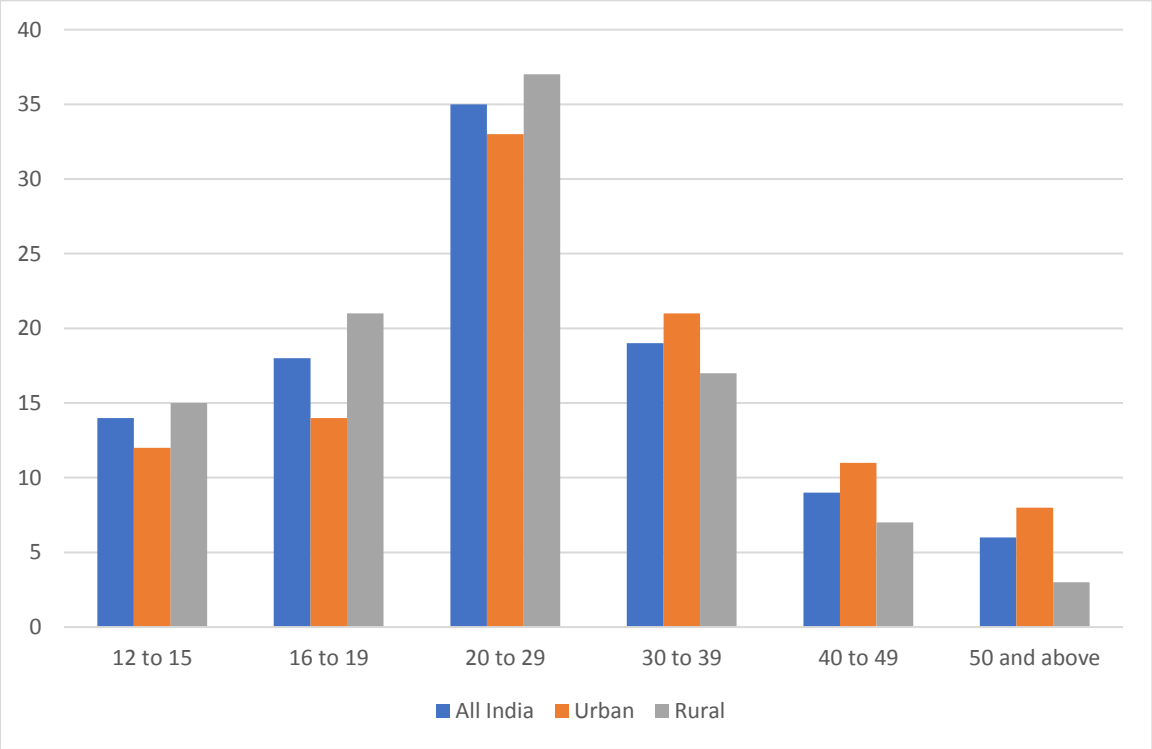
3 Ernst & Young (2013). *Reaping India's promised demographic dividend: industry in the driving seat.* EY and Federation of Indian Chambers of Commerce and Industry.

4 UNFPA (2019). *World Population Dashboard.* United Nations Population Fund.

bracket, leading to an increase in the working-age population from 761 million to 869 million between 2011 and 2020 (EY and FICCI, 2013)[5]. This phase of the demographic transition currently underway in India is expected to continue till around 2040 (World Bank 2012)[6].

The Internet in India is therefore, unsurprisingly, characterised by the striking youth of its user base. Two-thirds of Internet users in India are aged between 12 and 29 (Nielsen-IAMAI 2019), and most notably, as evident in Figure 1, a higher proportion of this youngest age group comprises rural users. India's rural young taking the pole position in Internet usage is an encouraging sign in the backdrop of the demographic dividend. It is this fact that forms the basis for the oft bandied about statement that rural India is the next engine for Internet growth. However, even as this is a promising statistic, its import ought to be tempered by the fact that around one in every five rural Internet users are among the section that use the Internet less often than once a week (ibid). Furthermore, nearly a third of rural users access the Internet between 15 to 30 minutes a day, while a similar proportion of urban users accesses it for more than 1 hour a day, which implies that improved rural connectivity, service quality, and affordability could potentially drive up daily rural Internet usage (ibid).

Figure 1: Age profile of internet users in India



Source: Nielsen-IAMAI (2019)

5 Ernst & Young (2013). *Reaping India's promised demographic dividend: industry in the driving seat.* EY and Federation of Indian Chambers of Commerce and Industry.

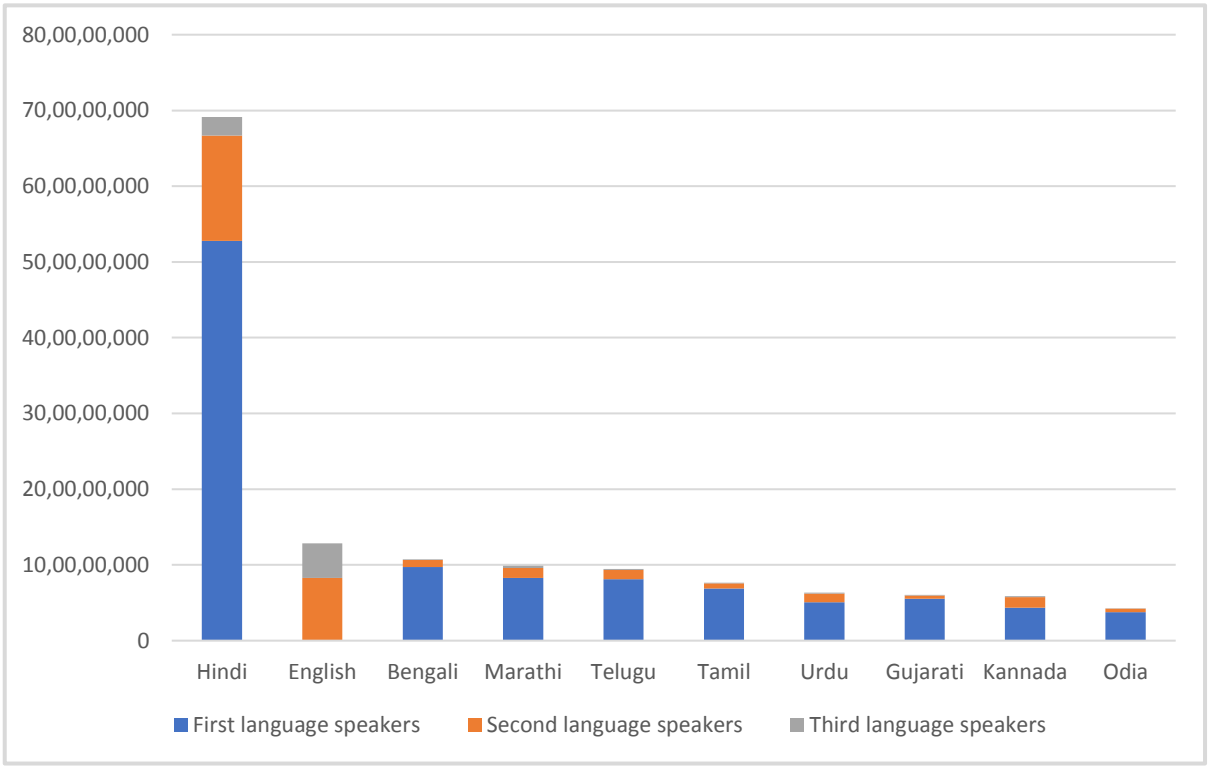
6 World Bank (2012). *More and Better Jobs in South Asia.* Washington DC: The International Bank for Reconstruction and Development, The World Bank.

2. The language barrier

Most Indians continue to live in rural areas and small towns, with different linguistic practices, and scanty knowledge of English. Each Indian state also happens to be pluri-cultural, besides showing a great degree of multilingualism (Census of India 2001 and 2011). It would not be unfair to say that English speakers very much constitute India's elite. Data from a 2019 survey by Lok Foundation and the University of Oxford, conducted by the Centre for Monitoring Indian Economy (CMIE) on the demographic profile of English language speakers in India suggests that the proportion of English speakers may also be shrinking (Livemint 2019). The 2011 Census showed that English is the primary language, or the mother tongue, of merely 256,000 people (virtually invisible as a result in Figure 2), the second language of 83 million people, and the third language of a further 46 million people. This means that English is the second-most widely spoken language after Hindi (which includes more than 50 dialects, for instance Bhojpuri, which is itself spoken by more than 50 million Indians).

A total of 528 million speak Hindi as a first language (Census of India 2011), making it both the most widely spoken first as well as second language in India. On the other hand, English is only the 44th most widely spoken first language, even though it is the second-most widely spoken second language. 6 percent of respondents to the Lok Foundation-Oxford University survey reported that they could speak English, which is less that what the 2011 census revealed. Aggregating for responses covering mother tongue, second and third language, the 2011 census records that more than 10 percent of Indians reported being able to speak at least some English.

Figure 2: India's top 10 most spoken languages



Source: Census of India (2011)

English is far more an urban than a rural phenomenon with only 3 percent of rural respondents reporting that they can speak English, as against 12 percent of urban respondents (CMIE 2019). One can also find a distinct class element at play in these dynamics. 41 percent of high-income respondents could speak English as against less than 2 percent of low-income respondents, while one-third of all graduates could speak English, indicating its strong linkage with being formally educated (ibid).

Indian language user base has increased steadily, from 42 million in 2011 (38 percent of the total Internet user base in India) to 234 million in 2015 (57 percent), and is projected to rise to 536 million in 2021, which would translate into 73 percent of the user base (KPMG 2017). Indian language Internet usage is being spurred on by reduced mobile data charges across operators, rising disposable incomes, growth in overall Internet penetration and smartphone users. Rural Indian language Internet users also have higher engagement levels, at nearly 530 minutes per week, compared to urban users at around 487 minutes per week (ibid). However, even though the gap in absolute numbers of Internet users is not significant between urban and rural users, the fact that approximately 70 percent of rural users do not actively use the Internet, and that the size of the overall rural populace far outweighs the urban, the cliché surrounding India's rural Internet potential wafts away. The numbers are self-evident, and the next vehicle of growth in usage will indeed be in the *mofussil*.

One of the most pertinent routes to ensure that the largest section of the population which is both rural and more comfortable with Indian languages, is through internationalised domain names (IDNs), which enable people around the world to use domain names in local languages and scripts. It must be emphasised that IDNs are among a variety of other interventions that will be able to effectively on-board rural populations, not least improving actual rural connectivity which requires infrastructural investments. Therefore, the scope of this paper is limited to assessing IDNs, their track record in enabling a multilingual Internet both globally and in India, as well as the opportunities and challenges they pose. Furthermore, even though this paper discusses the technical aspects of IDNs in broad terms, it does not explicitly delve into the purely technical domain, and restricts itself solely to technical inferences which could potentially aid policy formulation and course-correction.

3. An introduction to internationalised domain names

Internationalized Domain Names (IDNs) are domain names represented in local language characters. IDNs are formed using characters from different scripts, such as Arabic, Chinese, Cyrillic or Devanagari. These are encoded by the Unicode standard and used as allowed by relevant IDN protocols. The Internet Corporation for Assigned Names and Numbers (ICANN), has instituted the IDN Program to assist in the development and promotion of a multilingual Internet using IDNs. The program is primarily focused on the planning and implementation of IDN top-level domains (TLDs), including IDN country code TLDs and generic TLDs.

Interoperability and unique resolution are the building blocks of domain names' pivotal role in the Internet's addressing system. This essentially means that any user, anywhere in the world, who is connected to the Internet, can reach the same destination by typing in a domain name,

which can either be part of a web address or an email address). The bane of the issue in internationalising the Domain Name System (DNS) lies in a restricted character set within the Domain Name System: the American Standard Code for Information Interchange (ASCII), consisting of a to z, 0 to 9, and the hyphen. This was necessitated by technical constraints and the overriding priority of interoperability. Internationalised domain names can, however, contain letters or characters from non-ASCII scripts.

Technical standards to internationalise domain names were developed starting in the mid-1990s, and solution was devised which retained the DNS's restricted character set, while transliterating every other character into it. Each series of non-ASCII characters could now be transliterated into a string of ASCII characters prefixed with xn-- . These xn-- ASCII forms of the domain names are meaningful to name servers that resolve domain names and, as a result, users can view the meaningful, transliterated characters while navigating the Internet, whilst the underlying technical resolution of domain names remains unchanged. The algorithm used to transform a Unicode Label into an ASCII string is known as Punycode. This ASCII string is prefixed with "xn--" (ACE prefix) to create an "A-label" or ACE label (ASCII Compatible Encoding) that the domain name system understands.

The implementation of IDNs commenced in 2000 at the second level (under .com and .net) and 2001 (.jp). In the following decade, a number of country-code top level domains (ccTLDs) deployed IDNs, primarily supporting local language character sets. While some countries experimented with other strategies for internationalising domain names, the IDN technology proved the most successful. Despite the development of this solution, IDNs remain technically complex to implement, and several technical challenges abound, including how to handle variant characters, which are prevalent in Arabic and Chinese scripts. Another challenge is the user-experience, e.g. consistent representation in browsers and full functionality in emails – this is called 'universal acceptance'.

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 percent of Internet web pages; English, with 72 percent of web pages, was the dominant language online. Recent reports indicate that other languages are growing rapidly online. For example, by 2010, only 20 percent of Wikipedia articles were in English, and by June 2018 this had fallen to less than 12 percent 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 the World Report has demonstrated that IDNs are strongly linked to local content.

However, IDNs are not immune to potential abuse. Researchers based at Tsinghua University, and the University of Texas at Dallas, performed a comprehensive measurement study[7] using IDNs discovered from 56 TLD zone files. Through correlating data from auxiliary sources like WHOIS, passive DNS and URL blacklists, they found that 1.4 million IDNs were actively registered under over 700 registrars, and regions within East Asia have seen prominent

[7] https://personal.utdallas.edu/~shao/papers/liu_dsn18.pdf

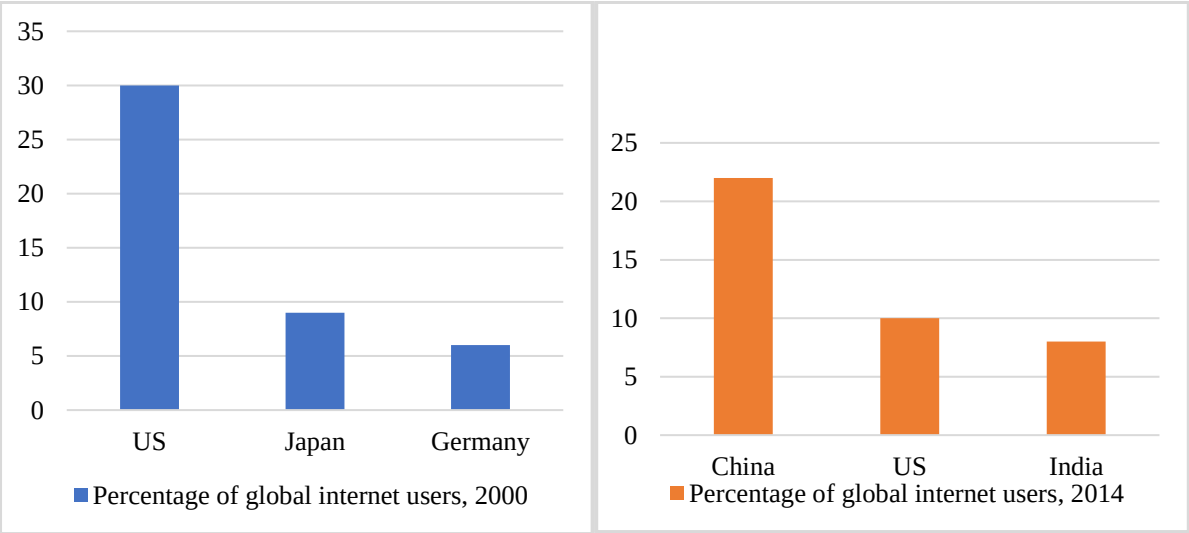
development in IDN registration. However, most of these registrations were opportunistic in that they were not currently associated with meaningful websites and had severe configuration issues (e.g., shared SSL certificates). They also note the rising trend of IDN abuse, with more than 6000 IDNs being determined malicious by URL blacklists, as well as thousands of IDNs showing high visual and semantic similarity to reputable brand domains (such as apple.com). Meanwhile, brand owners have only registered a few of these domains.

A cyber attacker can use the concept of homoglyphs (Unicode characters' sets resembling some ASCII characters, which are the basic building blocks for a domain name address) to spoof a domain name and lure an innocent user to visit a decoy domain instead of a legitimate one. IDN domain spoofing could be best detected at the end user side or by using a centralized monitoring solution. Egyptian researchers based at the Nile University in Cairo proposed a new centralized monitoring system[8] that can detect IDN spoofing attacks.

4. IDNs: a global overview

Over the period 2000-2014, the structure of global internet usage has undergone a sea change, with the largest number of internet users shifting from the developed to the developing world. According to ICANN (refer to Figure 3), in 2000, the US, Japan, and Germany were among the world's leading internet users, while in 2014, that space has largely been ceded to the emerging powers, China and India, which have seen a massive rise in Internet usage.[9] This is a dramatic reversal of pole positions, signalling that the Internet's next wave of growth will be in the largest developing economies.

Figure 3: Top three countries by percentage of total global internet users



Source: Internet Live Stats, in ICANN (2014)

8 https://docs.apwg.org/ecrimeresearch/2018/5359941.pdf

9 ICANN (2019). URL: https://www.icann.org/sites/default/files/assets/idn-access-domain-names-03sep15-en.pdf

Given that popular web platforms and applications are driving an ever-increasing support for multilingualism, IDNs have a vital role to play in enhancing linguistic diversity in cyberspace. For example, Facebook supports more than 110 languages (compared with 100 languages last year) and is actively increasing the languages that it supports, Google Translate is available for more than 100 languages, Twitter supports 34 languages. The world's most popular apps are also increasing the number of supported languages: WhatsApp is available in up to 60 languages, Instagram in 35 languages. The IDN market is more balanced in favour of emerging economies, and IDNs can be considered accurate predictors of the language of web content.[10]

Hybrid IDNs have been available at the second level with ASCII Top Level Domains (for example, παράδειγμα.eu) for nearly two decades. 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 based on non-Latin scripts (including Chinese, Arabic, and Hindi), the hybrid IDN/ASCII domains proved unsatisfactory. Right-to-left scripts, such as Arabic and Hebrew created bi-directional domain names when combined with left-to-right TLD extensions. This required users to be familiar with both their own language, and Latin scripts in order to navigate the Internet, which is an unsustainable model. Bi-directional domain names not only require Internet users to change the script while typing in a single web address, but also potentially confuse the strict hierarchy of the Domain Name System.

From 2006 onwards, Internet governance fora 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 Internet. From 2005, increasing pressure was placed on ICANN to implement IDNs in the root zone. Even as China and South Korea developed workarounds such as keyword searches at the domain name servers for .cn and .kr, where for those searching for domains within the country, the keyword system would resolve the domain without the user having to type the Latin-script domain ending (TLD), IDNs have undergone various stages of evolution since the turn of the century, illustrated in Table 1.

[10] Multilingualism and IDNs, IDN World Report (2018). URL: https://idnworldreport.eu/2018-2/multilingualism-and-idns/

Table 1: Evolution of IDNs

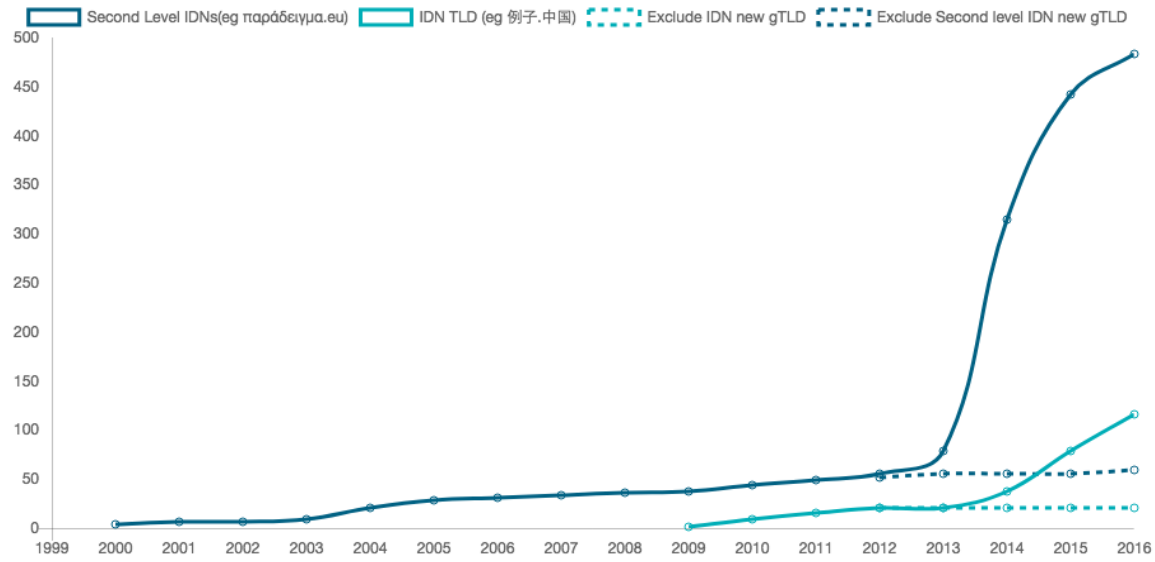
Year	Stage of Evolution
2000	Implementation of IDNs began in 2000 at the second level (under .com and .net) and 2001 (.jp), which was followed by several country code top-level domains (ccTLDs) deploying IDNs, primarily supporting local language character sets, over the course of the coming decade. This 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. 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.[11]
2005	Increasing pressure mounts on ICANN, the global coordinator of Internet domain names, to implement IDNs in the root zone.
2006	Internet governance discussions begin highlighting 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 Internet.
2007	Experimental .test IDN TLDs are introduced.
2009	ICANN Board approves a fast track process for IDN ccTLDs, describing the programme as a "top priority". Until late 2009, TLDs were restricted to only the Latin letters a to z without accents or symbols.
2010	ICANN took the historic step of approving ccTLDs in native scripts for four countries: مصر (Egypt), السعودية (Saudi Arabia), рф (Russian Federation) and امارات (United Arab Emirates).
2011	17 IDN ccTLDs launched in total. Following this, 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), سورية (Syrian Arab Republic), .қаз (Kazakhstan), срб (Serbia), 新加坡 and சிங்கப்பூர் (Singapore).
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 2015.
2017	More than 430 new gTLDs are now offering IDNs, including 45 IDN new gTLDs, and over 80 ccTLDs are offering IDNs (including more than 20 at the top level). Overall, more than 520 TLDs are now offering IDNs (at top and second level).

Source: Compiled from IDNs State of Play (2011) and ICANN (2019)

Figure 4 shows the cumulative number of IDN launches, both at the top and second levels, since 2000. The impact of ICANN's new gTLD programme is clearly noticeable, especially in the near-exponential upturn at the second level, witnessed 2013 onward.

[11] IDNs State of Play (2011)

Figure 4: Date of IDN deployment



Source: IDN World Report 2017

5. Countries and territories with their own IDN ccTLD

The IDN ccTLD Fast Track Process was launched on 16 November 2009 by ICANN. As of that date eligible countries and territories were able to request their respective IDN ccTLD(s) through the process. The process entails three steps, which include:

(i) Preparations in country/territory,

(ii) String Evaluation, and

(iii) String Delegation.

The following is the list of countries and territories that are free to enter Step 3, i.e. string delegation[12]:

[12] Icann.org. (2019). *IDN ccTLD Fast Track String Evaluation Completion - ICANN*. [online] Available at: https://www.icann.org/resources/pages/string-evaluation-completion-2014-02-19-en#u

Table 2: Country-wise IDNs and string delegation

ccTLD Reference	Country / Territory	Primary String	Desired Variant String(s)	String in English	Language	Script
DZ	Algeria	xn--lgbbat1ad8j الجزائر		Algeria / Al Jazair	Arabic	Arabic
AM	Armenia	xn--y9a3aq hшj		hye	Armenian	Armenian
BH	Bahrain	xn--mgbcpq6gpa1a البحرين		albahrain	Arabic	Arabic
BD	Bangladesh	xn--54b7fta0cc □□□□□		Bangla	Bangla	Bangla
BY	Belarus	xn--90ais бел		be	Belarusian, Russian	Cyrillic
BG	Bulgaria	xn--90ae бг		bg	Bulgarian	Cyrillic
CN	China	xn--fiqs8S , 中国 xn--fiqz9S , 中國		China	Chinese	Simplified Chinese, Traditional Chinese
EG	Egypt	xn--wgbh1c مصر		Egypt	Arabic	Arabic
EU	European Union	xn--e1a4c ею		eu	Bulgarian	Cyrillic
GE	Georgia	xn--node ბ ე		ge	Georgian	Georgian (Mkhedruli)
GR	Greece	xn--qxam ελ		el	Greek	Greek
HK	Hong Kong	xn--j6w193g 香港		Hong Kong	Chinese	Han (Simplified, Traditional)
	India	xn--h2brj9c □□□□		Bharat	Hindi	Devanagari
		xn--mgbbh1a71e بهارت		Bharat	Urdu	Arabic
		xn--fpcrj9c3d భారత్		Bharat	Telugu	Telugu
		xn--gecrj9c □□□□		Bharat	Gujarati	Gujarati
		xn--s9brj9c □□□□		Bharat	Punjabi	Gurmukhi
		xn--45brj9c □□□□		Bharat	Bengali	Bengali
		xn--xkc2dl3a5ee0h □□□□□□□		India	Tamil	Tamil
		xn--2scrj9c □□□□		Bharat	Kannada	Kannada

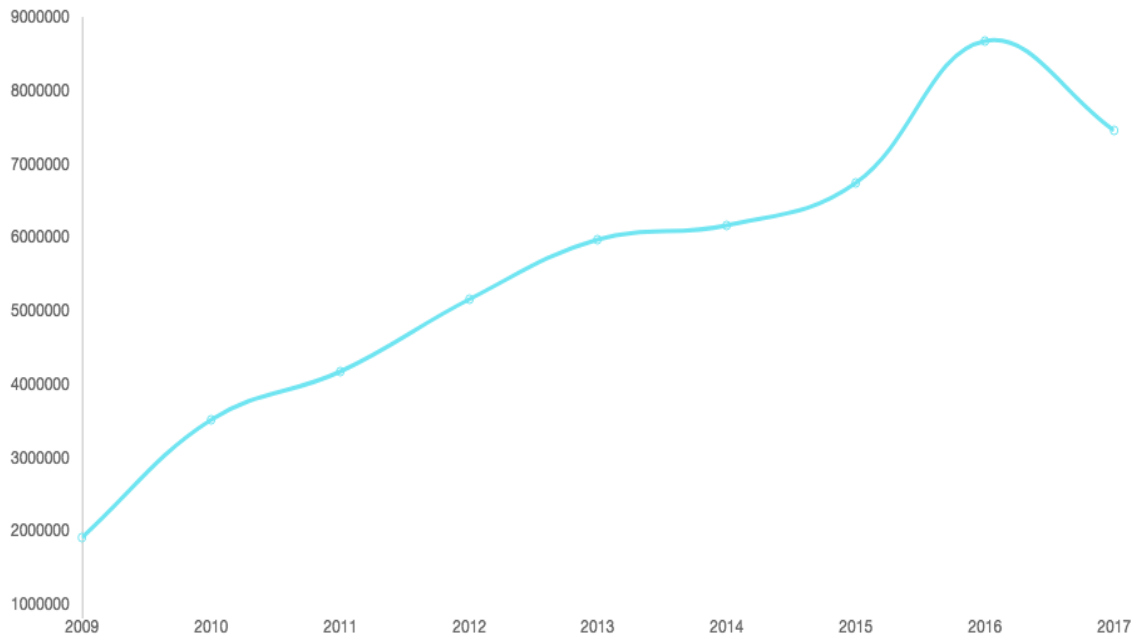
		xn--rvc1e0am3e ഭാരതം		Bharatam	Malayalam	Malayalam
		xn--45br5cyl □□□□		Bharat	Assamese	Bengali
		xn--3hcrj9c □□□□		Bharat	Oriya	Oriya
		xn--mgbbh1a بارت		Bharat	Kashmiri	Arabic
		xn--h2breg3eve □□□□□□		Bharatam	Sanskrit	Devanagari
		xn--h2brj9c8c □□□□□		Bharot	Santali	Devanagari
IN		xn--mgbgu82a پارت		Bharat	Sindhi	Arabic
IR	Iran, Islamic Republic of	xn--mgba3a4f16a ایران	xn--mgba3a4fra ایران	Iran	Persian	Arabic
IQ	Iraq	xn--mgbtx2b عراق		Iraq	Arabic	Arabic
JO	Jordan	xn--mgbayh7gpa الاردن		Al-Ordon	Arabic	Arabic
KZ	Kazakhstan	xn--80ao21a қаз		kaz	Kazakh	Cyrillic
KR	Korea, Republic of	xn--3e0b707e 한국		Republic of Korea	Korean	Hangul
MO	Macao	xn--mix891f 澳門	xn--mix082f 澳门	Macao	Chinese	Traditional Chinese
MK	Macedonia, The Former Yugoslav Republic of	xn--d1alf мкд		mkd	Macedonian	Cyrillic
MY	Malaysia	xn--mgbx4cd0ab .ملايسيا		Malaysia	Malay	Arabic
MR	Mauritania	xn--mgbah1a3hjkrd موريتانيا		Mauritania	Arabic	Arabic
MN	Mongolia	xn--l1acc мон		mon	Mongolian	Cyrillic
MA	Morocco	xn--mgbc0a9azcg المغرب		Morocco / al-Maghrib	Arabic	Arabic
OM	Oman	xn--mgb9awbf عمان		Oman	Arabic	Arabic
PK	Pakistan	xn--mgbai9azgqp6j پاکستان	xn--mgbai9a5eva00b پاکستان	Pakistan	Urdu	Arabic
PS	Palestinian Territory, Occupied	xn--ygbi2ammx فلسطين		Palestine	Arabic	Arabic
QA	Qatar	xn--wgbl6a قطر		Qatar	Arabic	Arabic
RU	Russian Federation	xn--p1ai рф		rf	Russian	Cyrillic

SA	Saudi Arabia	xn--mgberp4a5d4ar السعودية	xn-- mgberp4a5d4 a87g السعودية xn-- mgbqly7c0a6 7fbc السعودية xn-- mgbqly7cvafr السعوديه	AlSaudiah	Arabic	Arabic
RS	Serbia	xn--90a3ac срб		srb	Serbian	Cyrillic
SG	Singapore	xn--yfro4i67o 新加坡 xn-- clchc0ea0b2g2a9gcd □□□□□□□□□□□		Singapore	Chinese Tamil	Han Tamil
LK	Sri Lanka	xn--fzc2c9e2c □□□□ xn--xkc2al3hye2a □□□□□□		Lanka Ilangai	Sinhalese Tamil	Sinhala Tamil
SD	Sudan	xn--mgbpl2fh سودان		sudan	Arabic	Arabic
SY	Syrian Arab Republic	xn--ogbpf8fl سورية	xn--mgbtf8fl سوريا	Syria	Arabic	Arabic
TW	Taiwan	xn--kpry57d 台灣 xn--kprw13d 台湾	xn--nnx388a 臺灣	Taiwan	Chinese	Simplified Chinese, Traditional Chinese
TH	Thailand	xn--o3cw4h ไทย		Thai	Thai	Thai
TN	Tunisia	xn--pgbs0dh تونس		Tunis	Arabic	Arabic
UA	Ukraine	xn--j1amh укр		ukr	Ukrainian	Cyrillic
AE	United Arab Emirates	xn--mgbaam7a8h امارات		Emarat	Arabic	Arabic

6. Global IDN Trends

Since 2009, IDNs have largely grown. Figure 5 illustrates the total number of IDNs by year, and includes both top and second level IDN registrations (IDN World Report 2018).

Figure 5: Total IDNs by year since 2009

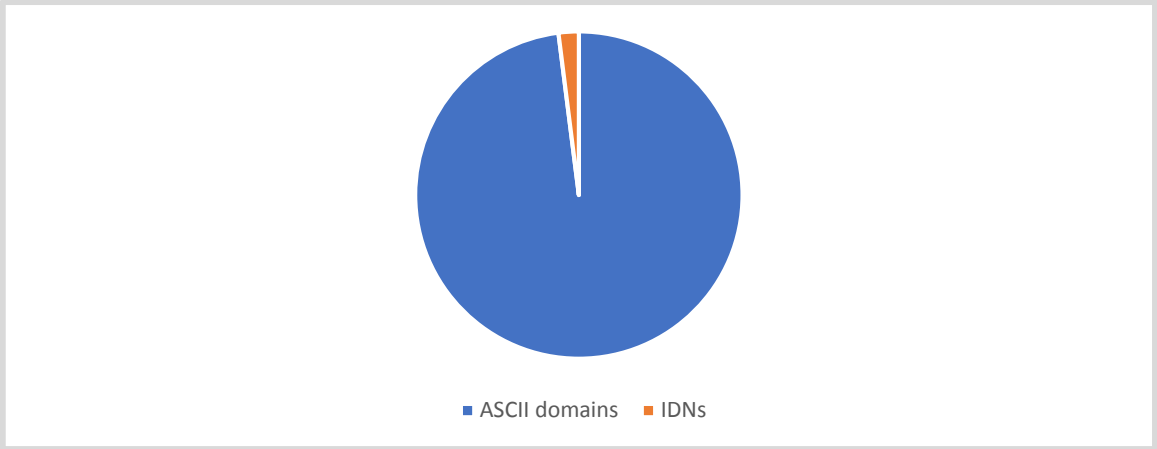


Source: IDN World Report (2018)

There were approximately 7.5 million IDNs as of December 2017 (IDN World Report 2018). The downturn, as evident in Figure 5, represents a drop in registrations of 15 percent during 2017, which was largely caused due to a change in policy by the Vietnam registry (.vn) which saw registrations at the second level under .vn fall from 977,000 in 2016 to 5,000 in 2017, coupled with a slight decline in registrations across the board (ibid). Excluding the impact of Vietnam, there was a drop of approximately 6 percent during 2017 (from 5.4 million to 5.1 million), compared to 28 percent between 2015-2016 and 9 percent growth in 2014-2015 (ibid).

As a percentage of the world's domains, IDNs comprise merely 2 percent (IDN World Report 2018). IDNs have consistently comprised approximately 2 percent of overall domain name registrations. In 2016, thanks to a growth in second level IDNs under the Chinese ccTLD, .cn, the percentage of overall domains was 3 percent

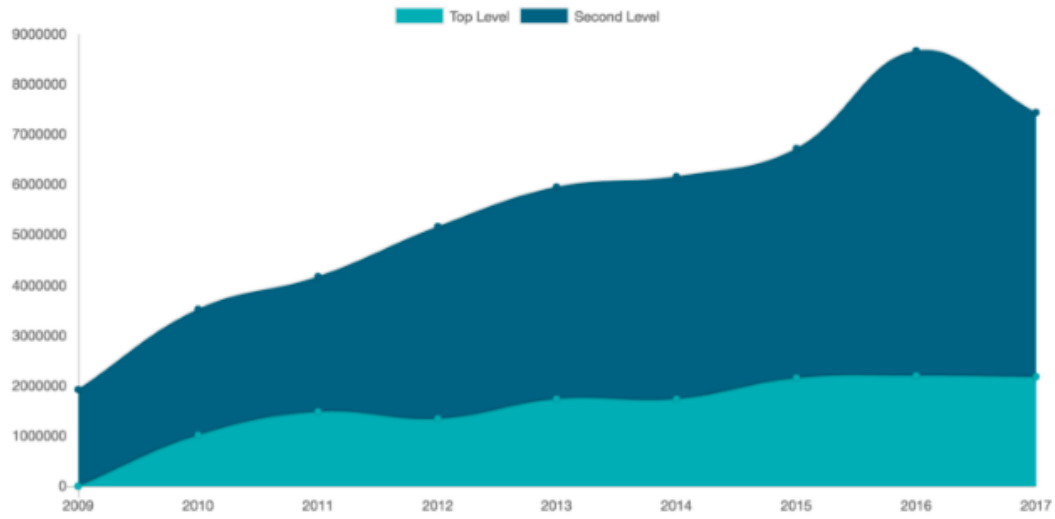
Figure 6: Global distribution of ASCII and non-ASCII domains



Source: IDN World Report (2018)

In December 2017, 71 percent of registered IDNs were at the second level, while 29 percent were at the top level, illustrated in Figure 7. During 2017, the reduction in second level IDNs under .vn contributed to a drop in the proportion of second level IDNs from 75 percent in December 2016. As of December 2017, there were 2.2 million top level IDNs (unchanged since December 2016) and 5.3 million second level IDNs (a reduction of 18 percent or 1.2 million since December 2016).

Figure 7: Global IDN distribution by levels



Source: IDN World Report (2018)

Meanwhile, the Chinese (Han) script TLDs dominate the list of top 10 top level IDN registries, while the Russian ccTLD, РФ continues to be the largest IDN top level domain, with almost 900,000 registrations. At the second level, three TLDs, .cn, .com, and .de have more than 500,000 registrations after which the numbers taper.

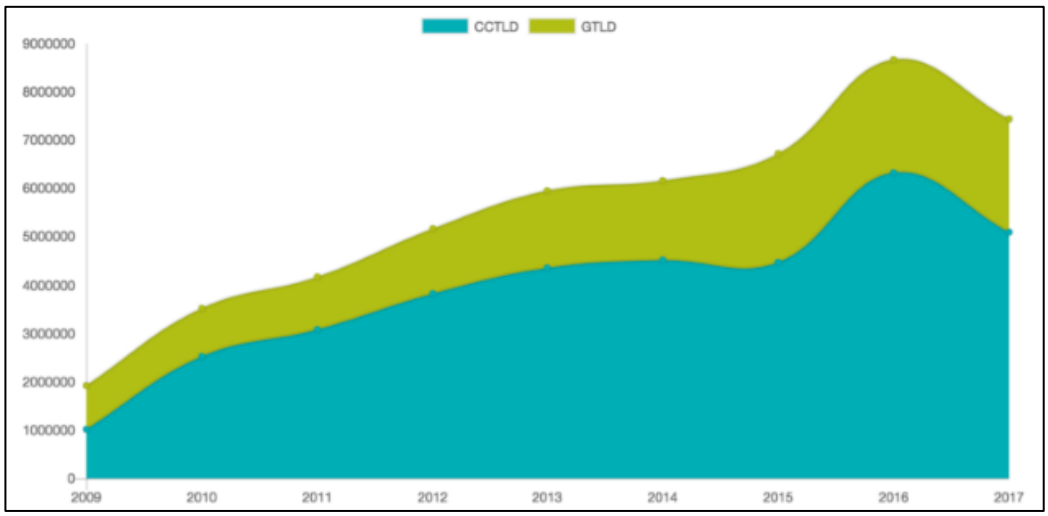
The top 20 IDN spaces have consistently comprised 90 percent of all IDN registrations. The following table illustrates the top 8 IDN spaces in the world, as of 2016 and 2017 (with the top 5 internationalised ccTLDs in bold):

Table 3: Global top 10 IDN spaces

IDN Space	Number of registrations in 2016	Number of registrations in 2017
.cn	2,100,000	2,100,000
.com	1,100,000	1,000,000
.vn	977,000	5,000
.рф	911,000	893,000
.de	635,000	648,000
.中国	474,000	474,000
.xyz	317,000	273,000
.net	272,000	225,000

Source: IDN World Report, 2017

Figure 8: Global IDNs by type



Source: IDN World Report, 2017

Furthermore, there are two types of Top-Level Domains:

a. **Country Code Top-Level Domains (ccTLD)**: assigned to countries or territories occurring in the ISO 3166 list. Each ccTLD is organised, managed and typically based at the local level in the country or territory designated by the TLD ending. Examples of ccTLDs include .eu, .de and .рф; and

b. **Generic Top-Level Domains (gTLD)**: do not necessarily have a geographical nexus and are coordinated at the international level through ICANN. The gTLD namespace expanded significantly in 2012, through the ICANN new gTLD programme. Examples of gTLDs include .com, .net and .网址.

In December 2017, there were 5.1 million IDNs (or 70 percent) in ccTLDs and 2.3 million IDNs (30 percent) in gTLDs, represented in Figure 8 above. Between December 2016-2017 the growth rate of IDNs in ccTLDs was -19 percent, while the growth rate of IDNs in gTLDs was 0 percent. All these figures include IDNs at both the top and second level.

Three scripts comprise 88 percent of all IDN registrations: Han, Latin and Cyrillic, with Han remaining the most widely used IDN script. Han and Cyrillic scripts represent more than 90 percent of top level IDNs. Cyrillic overtook Han in 2017 and now leads with 47 percent of top level IDN registrations. Over half of all second level domains are in the Han script, an ongoing effect of the 400 percent growth in IDNs under China's .cn through 2016. Second level Latin script IDNs fell from 40 percent to 29 percent over 2016-2017.

7. Roadblocks to IDN deployment and potential solutions

It is clear that the ability to use IDNs is a factor in the creation of local content. Local content is strongly related to the ability of web pages and applications to be customized to the requirements of a region or community. However, a key measure of the success of IDNs is an answer to the question: "Can they be used like other domain names?"

It is essential that IDNs behave just like other domain names and work, display and resolve correctly. This behavior can be referred to as "Universal Acceptance" or UA. Previous studies have shown that there are significant barriers to Universal Acceptance of IDNs. While there is progress, the pace of UA for IDNs is frustratingly sluggish. Progress toward UA for IDNs is especially slow in applications and security-related software. While there have been significant announcements of support for IDNs in email and other applications, the pace of uptake remains low.

Browsers continue to be a bright spot for the use, display and resolution of IDNs. Steady progress in browsers has been made in the last three years and the section that discusses IDNs and the World Wide Web notes promising improvements to the use of IDNs in browsers.

In addition to browsers, social media applications do a good job of displaying non-ASCII scripts and URLs (so-called 'linkification'). Still, no global social media service allows an individual to register a username that contains an IDN as part of the identifier. These small improvements cannot hide the fact that, in other parts of the Internet, Universal Acceptance is at best marginal and in some cases non-existent.

A major barrier to the uptake of IDNs is identified by European Registry of Internet Domain Names as lack of functionality on email and other applications. A survey conducted by EURid/CENTR found that 82 percent of participants highlighted the addition of email functionality as the single change that would improve IDN uptake. The basis for this view is that email is a key aspect of domain functionality and while it is unavailable for IDNs, their usefulness is limited. This is the also the case with IDN ccTLDs. Potential improvements could be "full support by the mobile environment" and the "ability to use IDNs in all applications including WhoIs and web browsers".

Further, user experience in registering an IDN, if cumbersome, is a significant disincentive to consider IDNs as a commercially viable option. The increasing reliance on search engines and portals vis-à-vis domain names as the principal navigation tools[13], reduce the necessity for memorable domain names, and may in turn impact on registration behaviour. This is even more so given that browser support for multiple scripts is improving and advances in search technology and predictive results are reducing the typing that a user has to perform.[14]

8. IDN issues being addressed by ICANN institutional structures

ICANN has instituted the IDN Program to assist in the development and promotion of a multilingual Internet using IDNs. The program is primarily focused on the planning and implementation of IDN top-level domains (TLDs), including IDN country code TLDs and generic TLDs. The IDN Program also supports projects geared towards effective use of IDNs at the second-level of the Domain Name System, as guided by the community. Further, ICANN has also issued guidelines, which serve as general standards for IDN registration policies and practices that are designed to minimize the risk of cybersquatting and consumer confusion, and to respect the interests of local languages and character sets. Registries seeking to deploy IDNs under their agreements with ICANN have been authorized to do so on the basis of the Guidelines.[15]

Communities using certain scripts, such as Arabic and Han, have identified that technically distinct Internationalized Domain Names (IDN) labels may be considered indistinguishable or interchangeable with other domain labels[16], and therefore regarded as the "same" domain labels, by the users of these communities, e.g. a label represented in simplified Chinese and equivalent traditional Chinese characters. The Board had resolved in 2010 for ICANN organization to look into the management mechanisms for IDN variant top-level domains (TLDs), to address relevant and complex linguistic, technical and policy issues. The subsequent work undertaken by ICANN organization and the community identified two challenges: (i) there is no accepted definition for variant TLDs, and (ii) there is no 'variant management' mechanism for TLDs. In July 2018, recommendations for managing IDN variant labels for TLDs were sought. The finalised recommendations will be presented to the ICANN Board, anticipated in March 2019.

Similarly, the Proposal for Sinhala Script Root Zone Label Generation Rules were posted for public comment in October 2018, soon followed by the formation of the Generation Panel to develop root zone Label Generation Rules (LGR) for the Hebrew script[17]. The public comment was released in February 2019. Meanwhile, a Study on Technical Use of Root Zone Label

---

13 Klensin, J. and Fältström, P. (2019). Internationalization of domain names; a history of technology development.

14 European Registry of Internet Domain Names (2011), Internationalised domain names: state of play, URL: https://unesdoc.unesco.org/ark:/48223/pf0000225888

15 ICANN (2019). URL: https://www.icann.org/resources/pages/implementation-guidelines-2012-02-25-en

16 https://www.icann.org/en/system/files/files/idn-variant-tld-exec-summary-25jan19-en.pdf

17 https://www.icann.org/news/announcement-2-2018-10-15-en

Generation Rules (RZ-LGR) was commissioned in August 2018, with an intent to make recommendations for the harmonized technical use of the RZ-LGR across IDN TLDs.

9. India and IDNs

To achieve a truly multilingual and accessible Internet, India, with its 22 constitutional languages, has an immediate need for IDNs to unlock the potential of the country's large non-English speaking population to be online. The complexity of this task is made starker given that, in addition to the constitutional languages, there are 122 languages with over 1,576 mother-tongues, most of which are written in 13 different scripts, with 53 more minor writing systems.[18] The Neo-Brahmi script Generation Panel (NBGP) has set the ball rolling towards this goal.

Further, the Neo-Brahmi script Generation Panel (NBGP)[19], formed by nine communities that use scripts derived from the Brahmi script, is developing Root Zone Label Generation Rules (LGR) for Bengali, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu scripts. NBGP intends to publish the proposals for the LGRs of these nine scripts in three sets, releasing proposals for scripts that share cross-script variant code points together to the extent possible, with proposals for Devanagari, Gurmukhi, and Gujarati Scripts' Root Zone Label Generation Rules being sought in July 2018, followed by Kannada, Oriya, and Telugu scripts in August 2018, and Malayalam and Tamil in September 2018.

According to the Universal Acceptance Steering Group (UASG), global internet servers are expected to be ready by June 2019 to enable people register complete website name in nine Indian scripts. LGR in root servers is intended to identify characters in Indian scripts, thereby allowing people to choose the complete name of a website at will. Eventual implementation of UA will enable people to use any name, word, or combination of letters in the different Indian scripts after ICANN accepts it and completes the process of allocating the system to various bodies across the globe that will facilitate registrations.[20]

Indian IDNs, transliterated as .bharat, in Hindi, Tamil, Bengali, Gujarati, Telugu and Urdu: .भारत, .□□□□□□□□, . ভারত, ભારત, భారత్, and تراهب respectively have been accepted by ICANN, however, Indic content accounts for barely 0.1 percent of the worldwide internet content.[21] 128 million users use Indic content. Of the 128 million users of Indic content, a majority (70 percent) of the users are only consuming video content. Lack of local language content online pushes this segment to use the internet for just video consumption.[22] The total share of Indian language internet users accessing digital payments is expected to increase from

---

18 Economic Times (2016). URL: https://tech.economictimes.indiatimes.com/news/internet/evolution-of-domain-name-system-for-indian-languages-is-under-way/52789166

19 https://www.icann.org/public-comments/devanagari-gurmukhi-gujarati-scripts-lgr-2018-07-27-en

20 Economic Times (2019). Internet servers may be ready by June to enable full website name in 9 Indian scripts. URL: https://economictimes.indiatimes.com/tech/internet/internet-servers-may-be-ready-by-june-to-enable-full-website-name-in-9-indian-scripts/articleshow/68670553.cms

21 IAMAI (March 2018). Available at: https://www.iamai.in/media/

22 IAMAI (August 2018). Available at: https://www.iamai.in/media/

28 percent today to 43 percent by 2021, thus limiting the potential market of an English only platform to 57 percent of user base.

More than 75 percent of Indian language internet users prefer mobile wallets over bank promoted websites and applications. Select leading mobile wallets have extended Indian language support on their platforms. Tamil, Telugu and Kannada users have higher potential for adoption of digital payments with more than half of their user base expected to avail the online service by 2021. The total Indian language internet users consuming digital news in the 8 Indian languages will exceed the English users by ~85 million in 2021, hence presenting a significantly larger market for regional languages. An average rural user spends 15 percent more time-consuming digital news than the urban counterpart.[23]

However, while these are encouraging indicators, we must be mindful of India's performance in IDN deployment vis-à-vis other countries. We select five representative countries, China, Russia, Germany, Japan, and South Korea, and compare India's IDN growth, year-on-year, with them in Table 4 below:

Table 4: Total IDNs, year-on-year

Country	2014	2015	2016	2017	2018
China	837,000	941,000	2,700,000	2,700,000	2,700,000
Russia	877,000	905,000	965,000	925,000	820,000
Germany	844,000	776,000	782,000	776,000	704,000
Japan	398,000	402,000	441,000	456,000	200,000
South Korea	194,000	178,000	185,000	181,000	668,000
India	**3,000**	**3,000**	**21,000**	**26,000**	**26,000**

Source: Compiled by the authors from the World Growth Map of IDNs (2019)

In any case, India was a late bloomer given that other countries such as Germany, Norway, Sweden and Portugal already had IDNs starting 2009. India's dismal track record is starkly visible, especially when viewed against China, which has steadily increased its IDN growth from 837,000 in 2014 to 2.7 million in 2018, compared to India's rise from 3,000 in 2014 to 26,000 in 2018.

The principal objectives of the Government of India's draft policy framework on Internationalised Domain Names in Indian Languages (2009) include the following[24]:

a. To ensure that Indian languages can have their rightful place in Internationalized Domain Names and that one can have a URL in an Indian language.

b. To initially permit such URLs in the following major languages/scripts: Devanagari (Marathi, Hindi, Konkani, Sanskrit and Nepali), Gujarati, Oriya, Punjabi, Malayalam, Tamil, Telugu, Kannada, Assamese, and Bangla and subsequently to be adapted for all

[23] KPMG (2017). Indian Languages – Defining India's Internet. Available at: https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf

[24] https://meity.gov.in/writereaddata/files/India-IDN-Policy.pdf

the 22 official languages including those using Perso-Arabic scripts: Urdu, Sindhi, Kashmiri.

c. To limit, at present the Indian language component to the Domain Name and localize the ccTLD, i.e. .in. To a large extent with some exceptions this will also allow language identification.

In August 2013, the Government of India released a policy framework for the .bharat ccTLD, which chiefly held that[25]:

a. Unlimited generic .भारत and कंपनी.भारत registration will be open to the public at 2nd level and 3rd level, except as provided herein.

b. The zones सरकार.भारत and शिक्षा.भारत will be reserved for the Government, and Educational institutions respectively. The registrations at the 3 level in these zones will be carried out by the Government, or an institution identified by the Government. Taking into account the Internet services offered in their respective sectors, .सरकार.भारत registration will be handled by NIC, शिक्षा.भारत by ERNET, as Registrars in these zones.

c. State Governments and Union Territories can register under .भारत category.

d. One character domain registration will not be allowed at the 2nd and 3rd levels for use by the general public.

e. Prior to the opening of registration at 2nd / 3rd levels, following category of names in the respective languages will be reserved: Constitutional Authorities, States/ Union Territories/ Cities, and Specific Names for Registry use

f. Trademark owners, registered companies and owners of intellectual property have a legitimate interest in protecting their brand. In the Internet domain, it is achieved by having a "Sunrise Period". A Sunrise period of 90 days from the opening of registration at second / third levels will be announced during which genuine registrations with proper verification will be allowed as per the policy for these registrations. Thereafter, the registrations will be open to public on first come first served basis.

g. The .IN Registry will have authority to deny or suspend the IDN registration to any one if it conflicts with the sovereign national interest or public order.

h. The .IN Registry may also open newly created generic top level zones, e.g. the Devanagari equivalent for offering registrations under these categories under IDN domain name.

[25] https://www.cdac.in/index.aspx?id=pdf_IDN_policy_framework

i. The Government of India will be the final authority for .भारत domain name policy. This will ensure its administration in accordance with the public interest and relevant laws of the country.

With the August 2013 policy framework and implementation plan in place, in August 2014, the Government of India launched the .bharat domain name in the Devanagari script covering several languages, including Hindi, Bodo, Dogri, Maithili, Marathi, Konkani, Nepali and Sindhi. Subsequently, the domain name was also launched in Bengali, Gujarati and Manipuri[26].

10. Conclusions and Policy Recommendations

The Government of India's efforts with the launch of bharat have failed to make any meaningful impact. Industry stakeholders hold that the .bharat IDN would not make any notable headway in rural India given the lack of critical information available in local languages (Livemint 2015). 24 percent urban and 20 percent rural Indian language Internet users are not aware of online government services, while more than 40 percent of urban and 30 percent of rural Indian language users are not interested in accessing government services online (KPMG 2017). Over 60 percent of rural Indian language Internet users stated that language was a barrier in accessing government services online (ibid). The lack of availability of both critical information and service delivery online in Indian languages calls into question the very purpose of providing Indic IDNs.

Following the failure of .bharat, the Government announced in 2016 that it would provide a free ".bharat" domain name for one year when a user buys a ".in" domain[27]. Even so, 2018 IDN statistics clearly reveal that this has also not made a difference. Overall, an analysis of these developments point to a classic case of a classic supply-driven failure of government intervention, where a good or service is made available without accounting for its existing demand, or lack thereof. Merely providing the bharat ccTLD, or pushing it as an add-on to .in registrations will not suffice if there exist no tangible incentive mechanisms for content creation in regional languages. In any case, once online, the website or domain name used does not make any significant difference to the user so long as useful information can be accessed.

The Indic language use of the Internet in India is not nearly as ubiquitous as Mandarin usage in China. Given that India has not proceeded on quite the same trajectory as China has in terms of local language content, it is difficult to reinvent the wheel at this point in time. Furthermore, IDN deployment in India comes at a time when the most popular modus for accessing online content is no longer through entering domain names in browsers. Increasing levels of e-commerce on social media, and the emergence of apps, has drastically altered the interaction model of Internet users. Domain names, therefore, do not hold the same pride of place they did a decade ago. The universal acceptance challenges facing IDNs, particularly with regard to email and applications, further exacerbate their uptake in India and around the world.

---

[26] https://www.thehindubusinessline.com/info-tech/now-get-internet-domain-names-in-regional-languages/article23159540.ece

[27] https://www.ciol.com/free-bharat-domain-name-with-in-purchase/

Verisign (2013) conducted qualitative research which led to five key insights regarding domain name extensions and IDNs:

1. IDN utility is not currently perceived as well appreciated
2. There is an initial resistance to adopting IDN.IDNs
3. Preferences do emerge for translation or transliteration dependent upon country
4. Interest in registering IDN.IDNs at this point appears moderate but there are reservations
5. Overall, registrar channel expectations for IDNs in general as well as specific IDN choices mirror those currently used in countries for finding and registering ASCII domain names.

All-English domain names currently dominate the Internet in India (Verisign 2013). Given the belief that English is the language for business in India, all-English domain names are viewed as easily understood, memorable, convenient, popular, and more global by respondents to the Verisign study. Respondents also stated they currently do not visit sites using IDNs, and while they know of some sites, especially newspaper sites, with native language content, the domain name is all in English. IDN.ASCII formats were found to be less popular than English, while Indian respondents were also found to not be familiar with them. They could not recall current websites using IDN.ASCII, perceived them to be difficult to input, and considered them to be applicable for local businesses only. Further, they aroused confusion over the rationale behind mixing two languages. On the other hand, while respondents were also not found to be familiar with IDN.IDN formats, they preferred them over IDN.ASCII formats since the former were not as confusing as mixing two languages. They believed IDN.IDN formats would provide better representation for companies, while there was also a perception that they could be relevant for personal use. However, they were also found to be difficult to input, and respondents felt, similar to IDN.ASCII, that they are more appropriate for local businesses.

While these perceptions are sobering, there is also a glimmer of hope in that particularly IDN.IDN formats are considered more representative for companies, even if only in comparison to IDN.ASCII formats. This signals that pure IDN formats could be seen favourably from both a commercial and an individual perspective, even though most Indians are unfamiliar with them Policy measures such as free bundling of .bharat with .in having not made the requisite impact, alternative approaches such as offering free .bharat IDNs to every new firm registering with the Registrar of Companies. This could potentially proliferate the uptake of IDNs by businesses that may not have considered an IDN previously. However, this alone would not be sufficient since, in isolation, this would remain a supply-side intervention with limited impact. Far greater emphasis needs to be paid to the creation of local language content by the Government, since both the existence and awareness of online government services in Indic languages is found lacking. Such content creation, which would require state initiative and funding in the absence of a viable incentive mechanism for markets, could be linked with IDNs, and would strongly signal to both users and businesses alike that the benefits of Indic languages online are aplenty.

About ICRIER

Established in August 1981, ICRIER is a policy-oriented, not-for-profit, economic policy think tank. ICRIER's main focus is to enhance the knowledge content of policy making by undertaking analytical research that is targeted at informing India's policy makers and also at improving the interface with the global economy.

ICRIER has two office locations in Delhi; in the institutional complex of India Habitat Centre and a new office at the Institutional Area, Sector 6, Pushp Vihar, New Delhi.

ICRIER's Board of Governors include leading academicians, policymakers, and representatives from the private sector. Mr. Pramod Bhasin is ICRIER's chairperson and Dr. Deepak Mishra is Director & Chief Executive.

ICRIER conducts thematic research in the following five thrust areas:

1. Growth, Employment and Macroeconomics (GEM)
2. Trade, Investment and External Relations (TIER)
3. Agriculture Policy, Sustainability and Innovation (APSI)
4. Digital Economy, Start-ups and Innovation (DESI)
5. Climate Change, Urbanization and Sustainability (CCUS)

To effectively disseminate research findings, ICRIER organises workshops, seminars and conferences to bring together academicians, policymakers, representatives from industry and media to create a more informed understanding on issues of major policy interest. ICRIER routinely invites distinguished scholars and policymakers from around the world to deliver public lectures and give seminars on economic themes of interest to contemporary India.