INTERNATIONALIZED DOMAIN NAMES-(IDN)

Inam Ullah,

Convener: Consortium for Language Advocacy & Networking (CLAN)

Member: Task Force on Arabic Internationalized Domain Names (TF-AIDN)

What are Internationalized Domain Names?

"Internationalized Domain Names (IDNs) are domain names represented by local language characters. Such domain names could contain letters or characters from non-ASCII scripts (for example, Arabic or Chinese)"

IDN TLD

An internationalized top-level domain (IDN TLD) is a top-level domain with a label that contains at least one character that is not a basic Latin letter (a,b,...z).

Domain names in Scripts other than Latin

παράδειγμα . δοκιμή

пример . испытание

דוגמא . בדיקה

실례 . 테스트

例子 . 測試

مثال . إختبار

例え . テスト

IDNs are Script-Based

- IDNs are Script-Based NOT Language-Based
- Each script may be used by many languages
- For example, Arabic script is used by
    - Arabic
    - Pashto
    - Persian
    - Sindhi
    - Urdu…just to mention a few

Major Scripts used by the languages of Asia, Europe and Africa

IDN TLDs introduced after 2009

What has changed with Top-Level Domains?

DOMAIN NAME
Latin letters a to z without accents or symbols.

.org .com .uk

2009

INTERNATIONALIZED DOMAIN NAMES ADDED
Characters from different scripts.

.عرب .中国 .рф

(.arab) (.china) (.rf)

Technical Demarcation: ASCII vs. Unicode

- ASCII is limited to English
- The Domain Name System (DNS) has not historically supported labels in other languages
- Other languages use letters and script outside this character set (a-z), (0-9) and (-)
- To make the DNS more accessible to users around the world
- The internet community has been working Internationalizing the Domain Names
- IDNs are essentially made up of Unicode characters

Why IDNs?

- Users prefer to use their native languages over English

- The combined population of China and Japan (1.5 billion) people – that's 5 times more than the U.S.

- Internet is global but you can't register a domain name and have URLs in your language

- IDN is the answer..

Significance of IDNs

Internationalized Domain Names allow people around the world to access domain names in their local languages.

# ICANN's IDN Program

Undertaking 28 + Scripts

Dealing with 19+ Generation Panels

There is a variation in the status of the Generation Panels.

Generation Panels' varying status
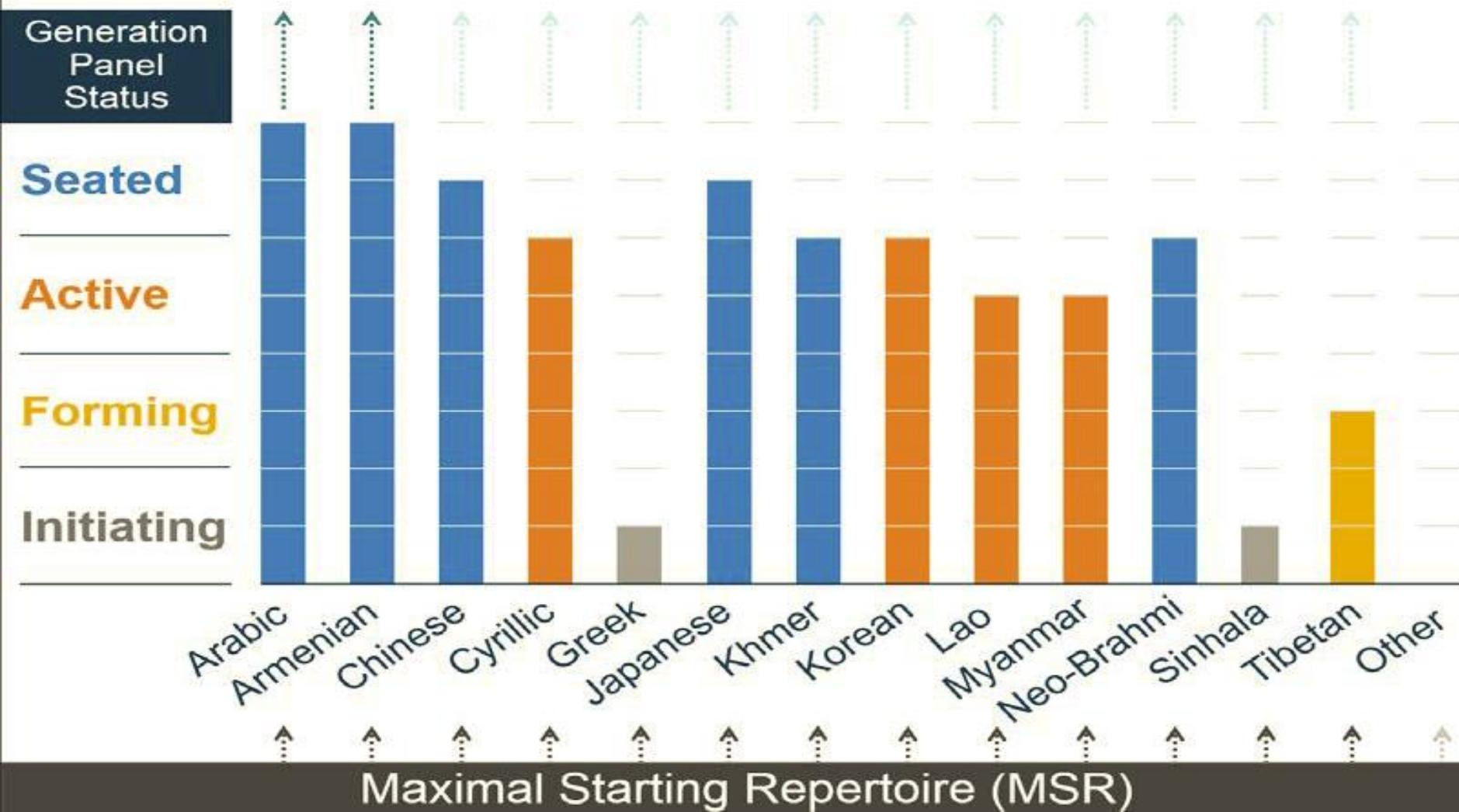
Generation Penals under IDN Program

- *Arabic*
- *Armenian*
- *Chinese*
- *Cyrillic*
- *Ethiopic*
- *Greek*
- *Japanese*
- *Khmer*
- *Korean*
- *Lao*
- *Neo-Brahimi*
- *Thai*
- *Tibetan*

IDN cc-TLDs

IDN Country Code Top-Level Domains

ccTLDs: 47/37*

укр (ua)

бел (by)

срб (rs)

мкд (mk)

ελ (gr)

المغرب (ma)

الجزائر (dz)

تونس (tn)

საქ (ge)

հայ (am)

سورية (sy)

فلسطين (ps)

مصر (eg)

سودان (sd)

الاردن (jo)

عراق (iq)

ايران (ir)

السعودية (sa)

اليمن (ye)

қаз (kz)

پاکستان (pk)

عمان (om)

امارات (ae)

قطر (qa)

भारत
بھارت
ಭಾರತ
ভাৰত
இந்தியா
ভারত
(in)

বাংলা (bd)

ලංකා
இலங்கை
(lk)

рф (ru)

Мон (mn)

中国
中國
(cn)

ไทย (th)

新加坡
சிங்கப்பூர்
(sg)

ملیسیا

한국 (kr)

台灣
台湾
(tw)

香港 (hk)

澳門 (mo)
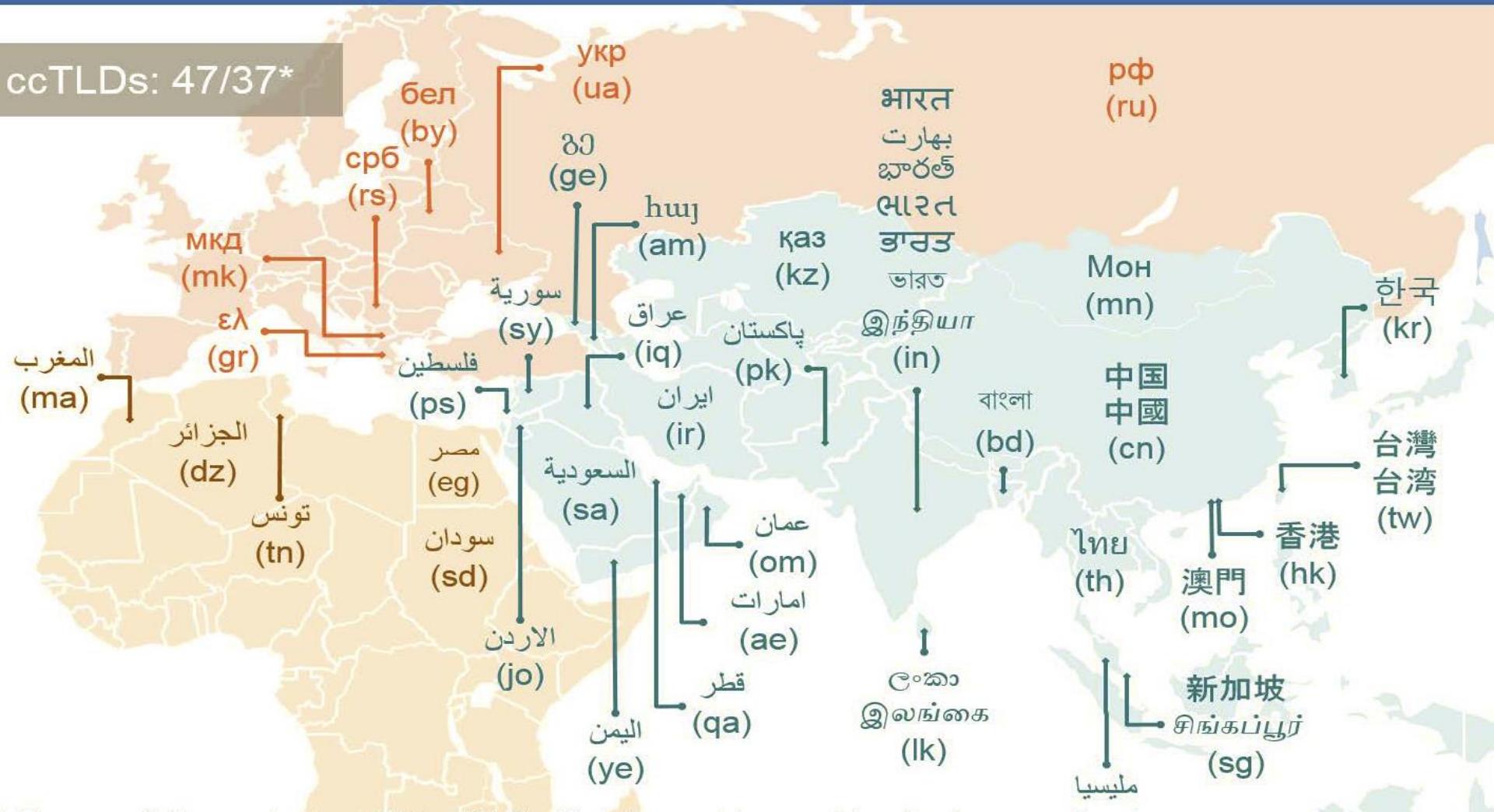
Arabic Script

Used by languages geographically distributed in the countries and regions of:

1. **Middle East**
2. **South and West Asia**
3. **North and Sub-Saharan Africa**
4. **South-East Asia**

Essentials for IDNs

- A set of Unicode Characters (Code Points)
- A set of specified variants

Major Languages Using Arabic Script

- Arabic
- Urdu
- Persian
- Pashto
- Uyghur
- Malay
- Hausa
- Wolof

TF-AIDN: Arabic GP

▫ Creation and oversight by community based Middle East Strategy Working Group (MESWG)

➢ 33 members from 18 countries

➢ Experts from a variety of backgrounds

Jobs TF-AIDN Accomplished

- Defining the code point principles
- Finalizing the code point repertoire
- Defining the code point variant principles
- Finalizing the code point variant sets
- Analyzing allocatable vs. blocked variants
- Finalizing code point variant types
- Analyzing Whole Label Evaluation rules
- Finalizing WLE rules
- Documenting the process
- Creating XML LGR for Arabic script LGR proposal
- Finalizing the Documentation and XML LGR proposal for submission to ICANN

Code Points undertook by TF-AIDN

Description	Number	Language / Example	Glyph
Total number of code points considered	343		
Code points PVALID according to IDNA 2008	289		
Code points included in MSR-1	239		
Code points for which TF-AIDN found evidence of use	167		
Specific code points shortlisted but excluded in IDNA 2008	1	Jawi / U+0662 –	٢
Specific code points shortlisted but excluded in MSR-1	5	Sindhi / U+06FD	۽
Code points where TF-AIDN found evidence of use but which are out of scope based on general MSR rules (Threatened or declining language, optional character)	32	Ormuri / U+076B Kurdish / U+0692	ڒ
Number of code points suggested for LGR by TF-AIDN	129		

Current Code Point Repertoire

Arabic Script TLDs Assigned or Delegated

الجزائر	بهارت	تونس
عمان	المغرب	مصر
ايران	السعودية	قطر
امارات	سودان	فلسطين
بازار	ماليسيا	عراق
پاكستان	شبكة	موقع
الاردن	سورية	

Situation of Pakistani Languages

Number of speakers of larger languages (Wikipedia)

Language	2008 estimate		Areas of Predominance
Punjabi	76,367,360	44.17%	Punjab
Pashto	26,692,890	15.44%	KP
Sindhi	24,410,910	14.12%	Rural Sindh
Saraiki	18,019,610	10.42%	Punjab
Urdu	13,120,540	7.59%	Urban Sindh and urban Pakistan
Balochi	6,204,540	3.59%	Balochistan

Scripts Used by Pakistani Languages

Almost all Pakistani languages use Arabic Script except:

Gujarati   (uses its own script)

Language Tables of Pakistani Languages for IDN

Language tables almost finalized:

Urdu & Pashto

Other Language Tables in progress:

Sindhi, Punjabi, Balochi, Balti, Shina, Seraiki,

Issues and Challenges

- Many characters can be confused with others
- Problem exists in ASCII as well
  - Digit "1" and lower-case "l"
  - Digit "0" and upper-case "O"
- IDNs increasing the character collection
  - From 64 in ASCII
  - To tens of thousands in Unicode
- This kind of confusion
  - create opportunities for user mistakes
  - and fraud