Internationalized Domain Names (IDNs) Updates

Pitinan Kooarmornpatana
IDN Programs Director

ICANN 81 Prep Week
31 October 2024

Agenda

- IDN Annual Report 2024
- IDN Implementation Guideline version 4.1
- IDN Table Reviewing process of IDN tables without Reference LGR
- Process to develop the additional Reference LGRs
- Community updates on additional Reference LGRs
  - Indonesian scripts
  - Thaana script
  - Languages in Unified Canadian Aboriginal Syllabics script
- Introduction of IDN topics at ICANN 81 Meeting Week
- Q&A

IDN Annual Report 2024

IDN Annual Reports

- Since 2021 ICANN org has published IDN Annual Report to provide an overview of the implementation of IDNs by ICANN org and the ICANN community.

- IDN Annual Report June 2024 includes:
  - Status of IDNs
  - IDN Policy Development
    - IDNs in the New gTLD Program: Next Round
    - GNSO EPDP on IDN
    - ccNSO ccPDP4
  - Status of IDN-Related Projects at ICANN
    - String Similarity Review
    - Single Character TLDs
    - Second-Level Reference Label Generation Rules
    - Label Generation Rule (LGR) Tool
    - IDN Implementation Guidelines
  - IDN Implementation and Operations at ICANN Org

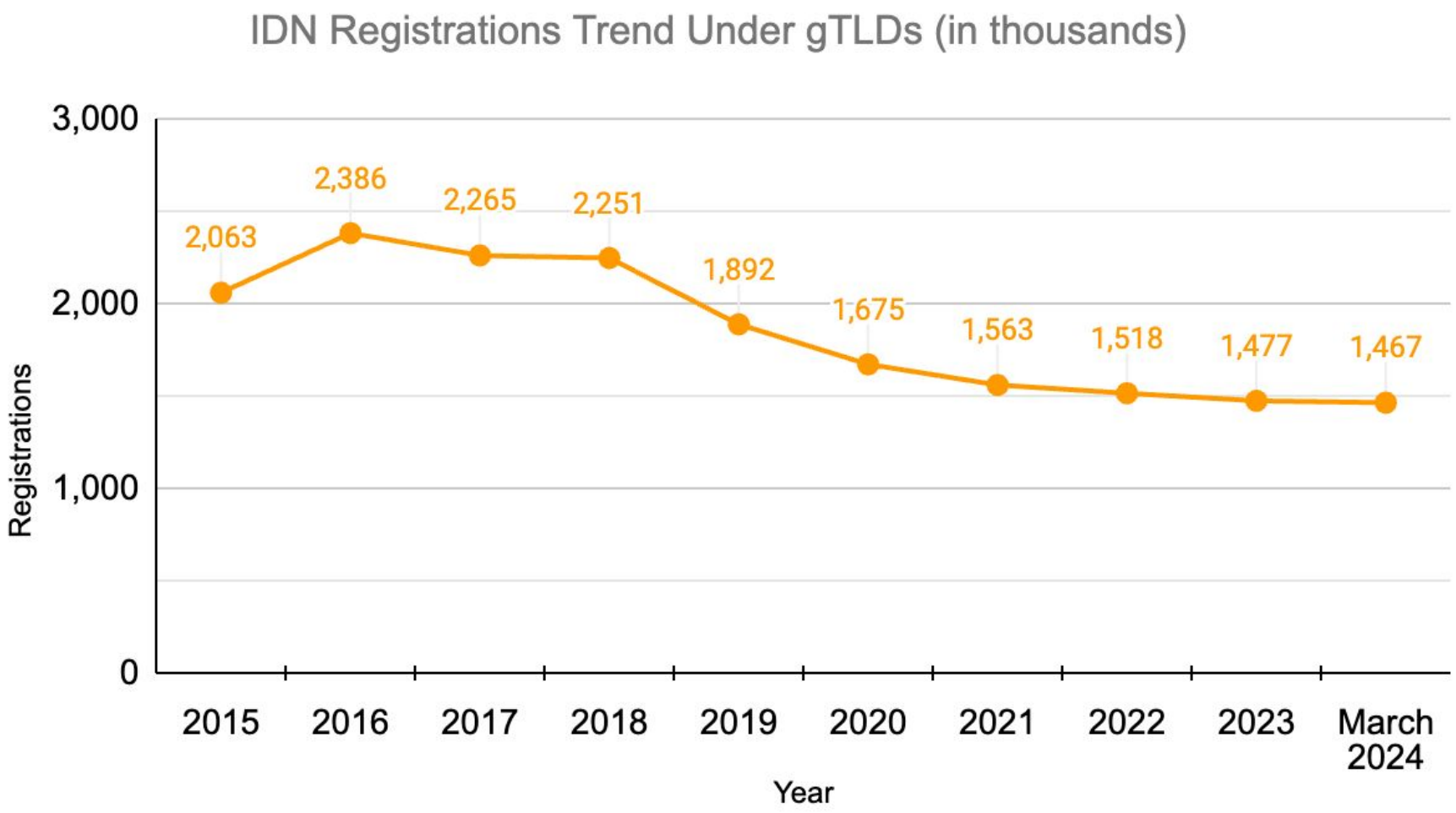
- Previous reports are available at https://www.icann.org/resources/pages/more-2012-05-08-en

Delegated IDNs at Top-Level (June 2024)

	ASCII TLDs	IDN TLDs	Total TLDs
ccTLD	248	61	309
gTLDs	1,048	90	1,138
Total	**1,296**	**151**	**1,447**

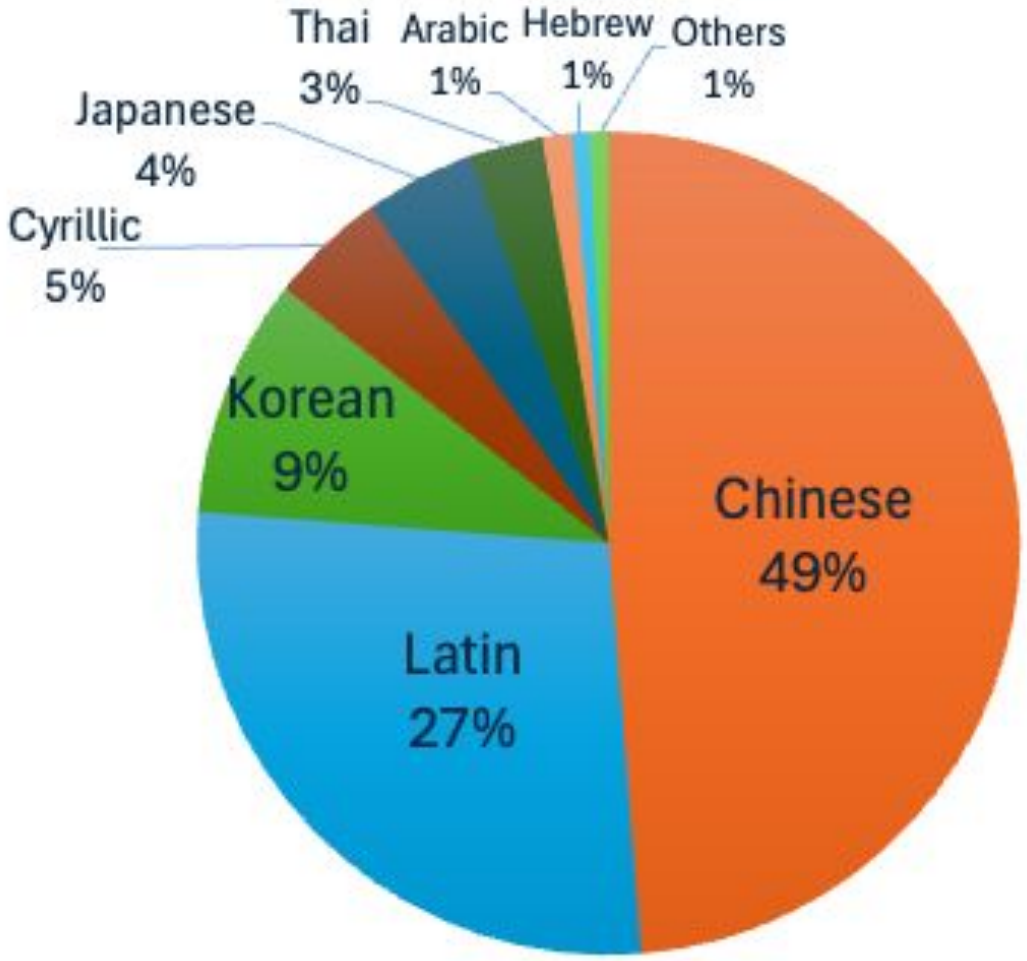
- The 151 IDN TLDs delegated to date represent 37 languages in 23 scripts.
  - The languages include Arabic, Armenian, Assamese, Bangla, Belarusian, Bengali, Bulgarian, Chinese, Georgian, Greek, Gujarati, Hebrew, Hindi, Japanese, Kannada, Kashmiri, Kazakh, Korean, Lao, Macedonian, Malay, Malayalam, Mongolian, Oriya, Persian, Punjabi, Russian, Sanskrit, Santali, Serbian, Sindhi, Sinhalese, Tamil, Telugu, Thai, Ukrainian, and Urdu.
  - The scripts include Arabic, Armenian, Bengali, Cyrillic, Devanagari, Georgian, Greek, Gujarati, Gurmukhi, Han, Hangul, Hebrew, Hiragana, Kannada, Katakana, Lao, Latin, Malayalam, Oriya, Sinhala, Tamil, Telugu, and Thai.
- Of these, 90 IDN gTLDs delegated, representing 10 languages in 10 scripts.

IDNs Registrations at Second-Level



- 1.467 million IDN registrations under gTLDs.
- IDN registrations under gTLDs have been decreasing over the past few years but at a slower rate.

IDN Registration under gTLDs by Script (March 2024)



- A total of 48.74 percent of IDN registrations are in the Chinese (Han) script, followed by 27.47 percent of registrations in the Latin script.

IDN Implementation Guideline version 4.1

Background

- IDN Implementation Guidelines relate to second-level registration policies and practices for IDNs.
- These contractually binding guidelines, which were developed by a community-led working group, are part of:
  - The Base Generic Top-Level Domain (gTLD) Registry Agreement.
  - The 2013 Registrar Accreditation Agreement.
- The guidelines are also a part of IDN country code top-level domain (ccTLD) Fast Track Process.
- The notice of implementation for version 3.0 was published in 2011.
- The ICANN Board adopted version 4.1 in September 2022, ICANN org published version 4.1 and has been working on its implementation.
- ICANN announced the implementation timeline in October 2024.

Summary of Preparation Phase

- The major updates in IDN Guidelines 4.1 include:
  - A strict compliance with IDNA2008 standard.
  - The implementation of secure and stable IDN tables to prevent whole-script confusable domain names with a reference to the available of Second Level Reference Label Generation Rules.
  - A requirement of making the applicable transition plans available.

- To gauge the significance of the impact of these Guidelines on contracted parties, ICANN org conducted a review in March 2024. ICANN org found that only 0.0003% of all registrations under gTLDs are potentially non-compliant with IDN Guidelines 4.1.

- The IDN Guidelines 4.1 implementation plan has been presented at ICANN public meetings and Contracted Parties Summits.

Implementation Timeline

- **Implementation Preparation - 31 October 2024 - 31 January 2025:**
  - Registry operators must continue to implement the IDN Guidelines 3.0 and should use this period to prepare for implementation of the IDN Guidelines 4.1.

- **Implementation Transition - 1 February 2025 - 30 April 2025:**
  - Registry operators may continue to implement the IDN Guidelines 3.0, or the IDN Guidelines 4.1, or elements of both.

- **IDN Guideline 4.1 Implementation in effect - 30 April 2025 and onward:**
  - The IDN Guideline 4.1 is effective and registry operators must comply in full with the IDN Guidelines 4.1 no later than 30 April 2025.

IDN Table Reviewing process of IDN tables without Reference LGR

Background

- IDN tables define which labels can be registered for a particular language or script under a TLD.
- IDN Tables are reviewed by ICANN for security and stability considerations during Registry Services Evaluation Policy (RSEP) IDN Services, Registry Service Provider (RSP) Change Testing or Pre-Delegation Testing (PDT).
- ICANN org continues to maintain the consistency and transparency of the IDN table review process, as requested by the Registry Stakeholder Group (RySG), by using
  - the Reference Label Generation Rules as a baseline, and
  - the LGR Tool to conduct the review.
- ICANN org published 25 script-based Reference LGRs while there are 168 scripts in the Unicode (version 16.0). Many scripts are not yet covered by the current set of reference LGRs.
- The Proposed Process to Review gTLD IDN Tables Without Reference Label Generation Rules was developed, in consultation with RySG, and published for public comment in August 2024.

Summary

- The Proposed Process to Review gTLD IDN Tables Without Reference Label Generation Rules proposed that during the IDN services or RST Process, if an IDN table review request is submitted before the relevant reference LGR or community input become available:
  - ICANN org will not be able to conduct the review and will request the requester to consider cancelling the request for that IDN table. The IDN table can be added in the future when the reference LGRs become available.
  - ICANN org will request RO to provide contact information of script experts and continue to develop additional reference LGRs based on community input.

- The public comment proceeding closed and summary report was published on 3 October 2024. The comments received do not provide any opposing views to the proposed process.

- ICANN org will proceed with the implementation of the Proposed Process to Review gTLD IDN Tables Without Reference Label Generation Rules.

Process to develop the additional Reference LGRs

Coverage of Reference LGRs

- 31 language-based LGRs
  - Arabic, Belarusian, Bosnian (Cyrillic), Bosnian (Latin), Bulgarian, Chinese, Danish, English, Finnish, French, German, Hebrew, Hindi, Hungarian, Icelandic, Italian, Japanese, Korean, Latvian, Lithuanian, Macedonian, Montenegrin, Norwegian, Polish, Portuguese, Russian, Serbian, Spanish, Swedish, Thai, and Ukrainian.
    - When a specific language-based LGR is not available, the IDN table can be reviewed by using a relevant script-based reference LGR. For example, a Dutch language IDN table can be reviewed by using the Latin script reference LGR.

- 25 script-based LGRs
  - Arabic, Armenian, Bangla (Bengali), Chinese, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Hebrew, Japanese (Hiragana, Katakana, Kanji [Han]), Kannada, Khmer, Korean (Hangul, Hanja [Han]), Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, and Thai.
    - These cover more than 386 languages which are actively used across the world

- ICANN continues developing additional reference LGRs with the script communities.

Scope of Work for Developing the LGR

- The language or script community proposes to form the reference LGR panel which includes linguistic experts, script encoding technical experts, and script users.

- The panel is tasked to analyze and determine three main parts of the LGR.
  - Repertoire
    - Goal is to create a mnemonic system for use in the Domain Name System (DNS) based on the https://www.rfc-editor.org/rfc/rfc6912.html.
    - Not required to completely cover a language or a script, may not form labels which are words in a language.
  - Variants
    - Goal is to define code point or code point sequences causing two (or more) labels functionally “the same” in a script.
    - Conservatism requires: maximizing “blocked” variants, minimize “allocatable” variants.
  - Rules
    - Goal is to reduce label space to prevent labels which should not be possible for various. reasons e.g. causing security issues, not licensed by the script.

Collaboration on Developing the Reference LGRs

- Expertise provided by language or script panel.
  - Prioritizing the language or script in the scope of work.
  - Word list or word corpus for each language or script.
  - Linguistic expertise to define repertoire, variants, rules.
  - Review the LGR and the test result.

- Expertise provided by ICANN org.
  - Technical support on developing the LGR in RFC 7940 format for panel to review.
  - Testing and provides testing results for community to review to see accuracy of ref LGR.
  - Taking an LGR through the public comment process before finalizing and publish an LGR.

- Depends on the time required by the panel, generally takes 6 – 12 months.

- Currently active panels includes.
  - Panel for multiple scripts used in Indonesia.
  - Panel for Thaana script.
  - Panel for multiple languages using Unified Canadian Aboriginal Syllabics script.

Steps in Developing a Reference LGR

Step	Task	Owner
1	Form the reference LGR panel	ICANN org, Panel
2	Develop the repertoire based on the usage data	Panel
3	Develop the variant, if applicable	Panel
4	Develop the rules, if applicable	Panel
5	Document discussion and decisions	Panel
6	Develop test labels	Panel
7	Generate the LGR in RFC 7940 format	ICANN org
8	Review the test result for see accuracy of reference LGR	Panel
9	Publish for the proposed reference LGR for public comment	ICANN org
10	Finalize the reference LGR after the public comment	Panel
11	Publish the final reference LGR at the ICANN webpage	ICANN org

Community updates on additional Reference LGRs

- Indonesian scripts
- Thaana script
- Languages in Unified Canadian Aboriginal Syllabics script

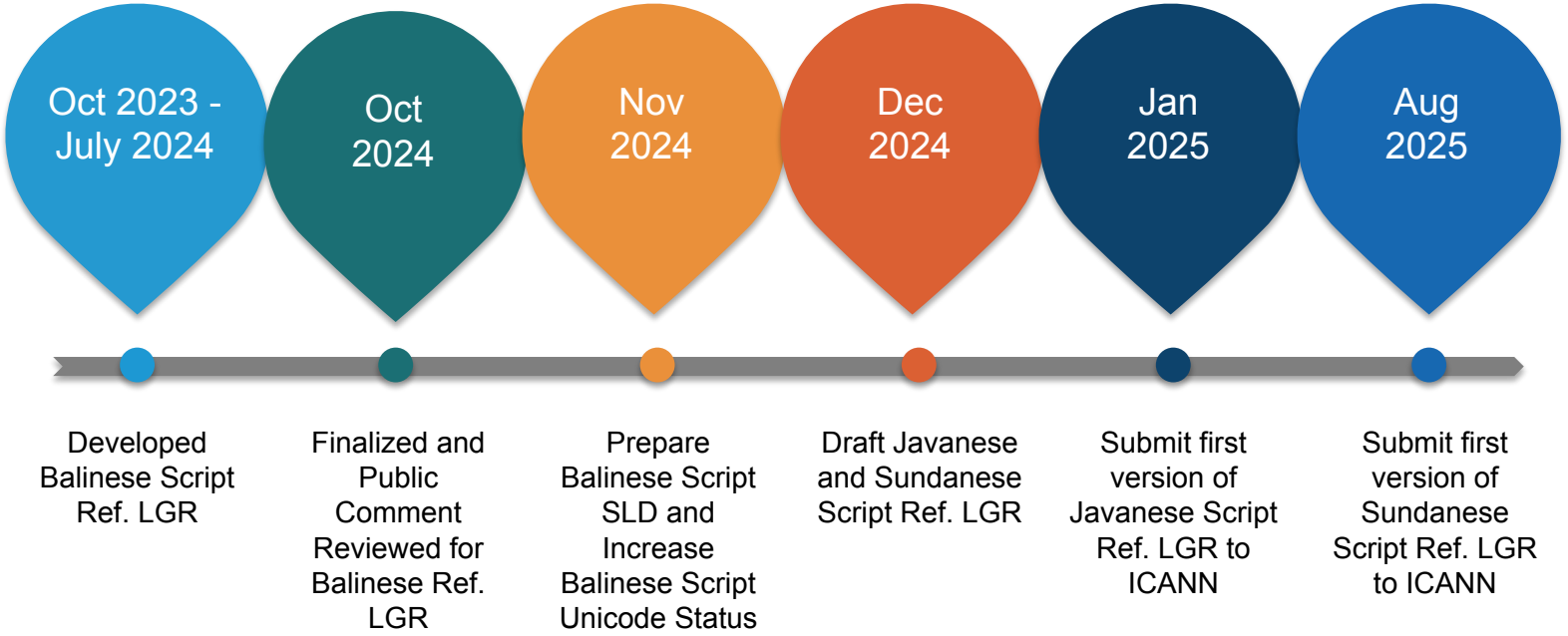
Reference LGR for Scripts used in Indonesia

Yudho Giri Sucahyo, Ph.D
Ilham Nurwansah, M.Ed

Background and Current Status

- Some of PANDI's efforts that have been made towards the IDN goal:
  - Collaborated with **UNESCO** and worked with more than **30 cultural institutions** in Indonesia, consisting of elements of universities, government, and the script and cultural communities.
  - Organized more than **40 activities** (competitions, symposiums, seminars, webinars, congresses) to encourage the **standardization** of the archipelago's script and its **massive** use in the **digital realm**.
  - Worked together with the **National Standardization Agency** (BSN) to obtain: Establishment of SNI FONT and KEYBOARD for **Javanese, Sundanese, Balinese, Kawi**, and **Pegon** Script.
  - Improving the status of the Indonesian scripts in UNICODE from **Limited Use** to **Recommended** by optimizing the digitization of scripts.
  - Working together with script **experts** especially in **Balinese**, **Javanese**, **Pegon,** and **Sundanese** which directly directed by ICANN to make a **Label Generation Rules (LGR)** proposal for the **second level domain** (SLD).

Next Steps



Publish Javanese and Sundanese Reference LGR

To Summarize

PANDI is committed to fully supporting this initiative and will continue to play a key role in the process, ensuring that these developments align with the broader goal of advancing Indonesian script towards Internationalized Domain Names (IDN). This commitment reflects PANDI's dedication to preserving Indonesia's rich linguistic heritage while promoting its digital presence globally.

Background and Current Status

- Unicode version 16.0.0 (last updated September 2, 2024)

UNICODE	**9 scripts**	**Javanese, Balinese, Sundanese, Makassarese, Rejang, Bataknese, Buginese, Kawi, Pegon (Arabic)**

Breakdown Status:

EXCLUDED SCRIPTS	4 scripts	Makassarese, Rejang, Buginese, Kawi
LIMITED USE SCRIPTS	4 scripts	Javanese, Balinese, Sundanese, Bataknese 
RECOMMENDED SCRIPTS	1 script	Pegon (Arabic extended-A & extended-C)

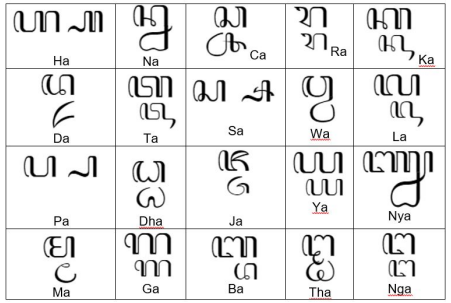
Background and Current Status

- Balinese script Reference LGR was developed
  - Online and on-site meetings October 2023 – July 2024.
  - Published for public for public comment, closed 24 October 2024.
  - Final version after public comment available at https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en.

- Other scripts' Progress:
  - Javanese: Developing LGR and collecting label set (+120,000 set).
  - Sundanese: Drafting LGR and collecting label set (+80,000 set).
  - Pegon: Drafting LGR and identifying label set for various Indonesian local languages (Sundanese, Javanese, Madurese, etc.; +131.000 set).

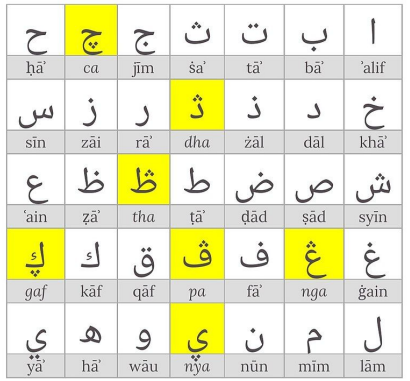
Challenges and Uniqueness

- **Javanese Script:**
  - A Brahmic script type with 91 characters in Unicode and thousands of syllable combinations.
  - It requires careful font rendering due to subjoined, conjoined, and vocalized (diacritic) features.
  - Some characters are visually similar, requiring alternatives or reduction in design.



- **Sundanese Script:**
  - A revitalized script (Brahmic type), now officially used in schools and public spaces.
  - Simplified characters reduce ambiguity, and it only uses a single line without subjoins.
  - Suitable for modern usage, especially in address bar display with vocalization diacritics.

- **Pegon Script:**
  - Part of the Arabic script family, making it eligible in the Root Zone.
  - Used in multiple languages with slightly different writing conventions.
  - Pegon employs two modes: with or without vocalization (diacritics).



Reference LGR for Thaana script

Naail Abdul Rahman
Co-Chair, Thaana Script Reference LGR Panel

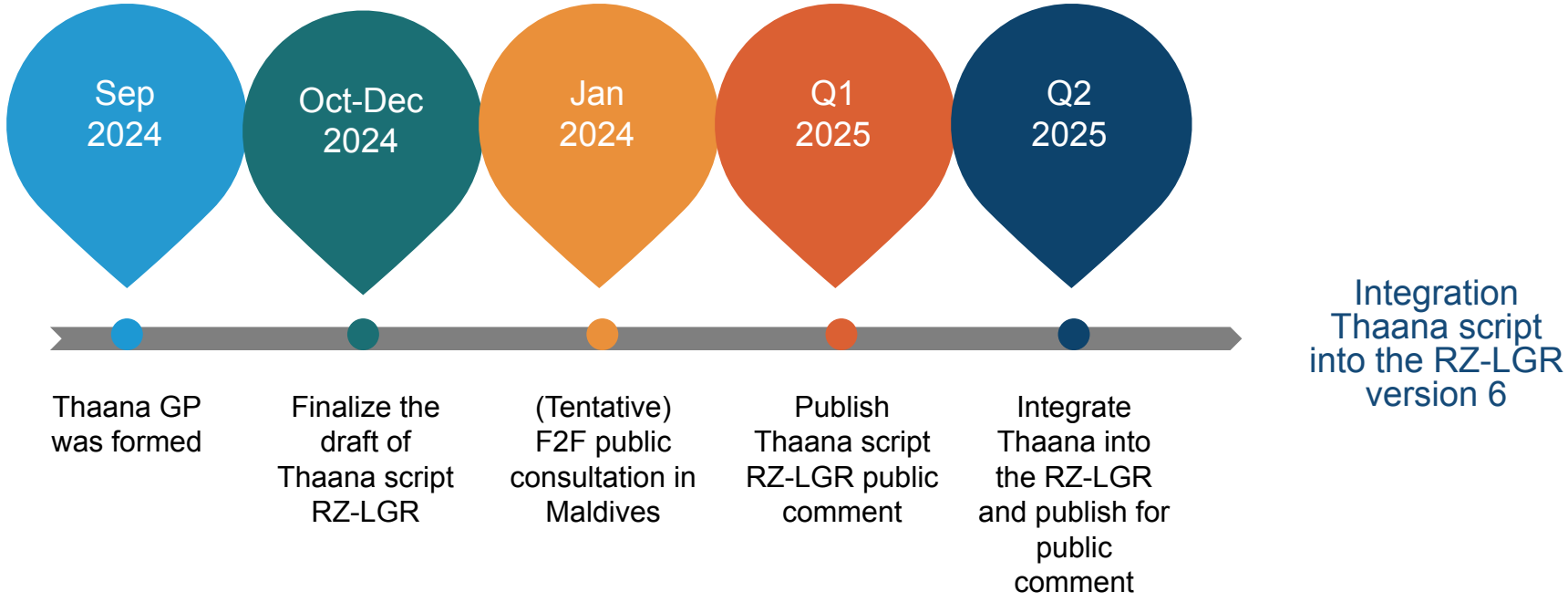
Background and Current Status

- Script information
  - ISO 15924 Code: Thaa, ISO 15924 Key N°: 170, ISO 15924 English Name: Thaana, Latin transliteration of native script name: Thaana, Native name of the script: ދިވެހި,
  - Geography: South Asian group of 22 atolls, located south-west of India
  - Population: 515, 122 (382,751 Maldivian, 132, 371 Expatriates) based on the recent official census (National Bureau of Statistics, 2023)
  - Earliest Existing records of Thaana dates back to 1705 AD, (Mohamed, 1999, p. 32)
  - Thaana has 25 consonants and 11 vowels.

- Thaana script Reference LGR was developed
  - Online and on-site meetings January – August 2024
  - Published for public for public comment, closed 24 October 2024
  - Final version after public comment available at https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en

Next Steps

- Continue to develop the Thaana Script Root Zone LGR
  - Thaana GP for developing Root Zone LGR was formed September 2024
  - RZ-LGR will be developed based on the LGR Procedure and the information available in the Reference LGR for the Second-Level
  - The Thaana Script RZ-LGR is planned to publish for public comment in Q1 2025



Root Zone LGR Panel Members

#	Name	Role	Designation	Organization	Expertise
1	Naail Abdul Rahman	Chair	Managing Partner, Software Engineer	Intellea Partnership	Unicode standard, Software Engineer
2	Aminath Saeed	Co-Chair	Senior Lecturer	The Maldives National University	Linguistics
3	Ali Fawaz Shareef	Member	Deputy Vice-Chancellor (Academic Affairs)	The Maldives National University	Academic
4	Abdulla Shafeeu (Abo)	Member	Co-founder	Akuru Type Marketplace	Type Designer
5	Hassan Irfan	Member	Deputy Minister of Cities	Local Government and Public Works.	Script user

Repertoire

Thaana[1][2]

Official Unicode Consortium code chart (PDF)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+078x	ހ	ށ	ނ	ރ	ބ	ޅ	ކ	އ	ވ	މ	ފ	ދ	ތ	ލ	ގ	ޏ
U+079x	ސ	ޑ	ޒ	ޓ	ޔ	ޕ	ޖ	ޗ	ޘ	ޙ	ޚ	ޛ	ޜ	ޝ	ޞ	ޟ
U+07Ax	ޠ	ޡ	ޢ	ޣ	ޤ	ޥ	ަ	ާ	ި	ީ	ު	ޫ	ެ	ޭ	ޮ	ޯ
U+07Bx	ް	ޱ														

consonants vowels extended consonants (excluded)

Reference LGR for Languages in Unified Canadian Aboriginal Syllabics (UCAS) Script

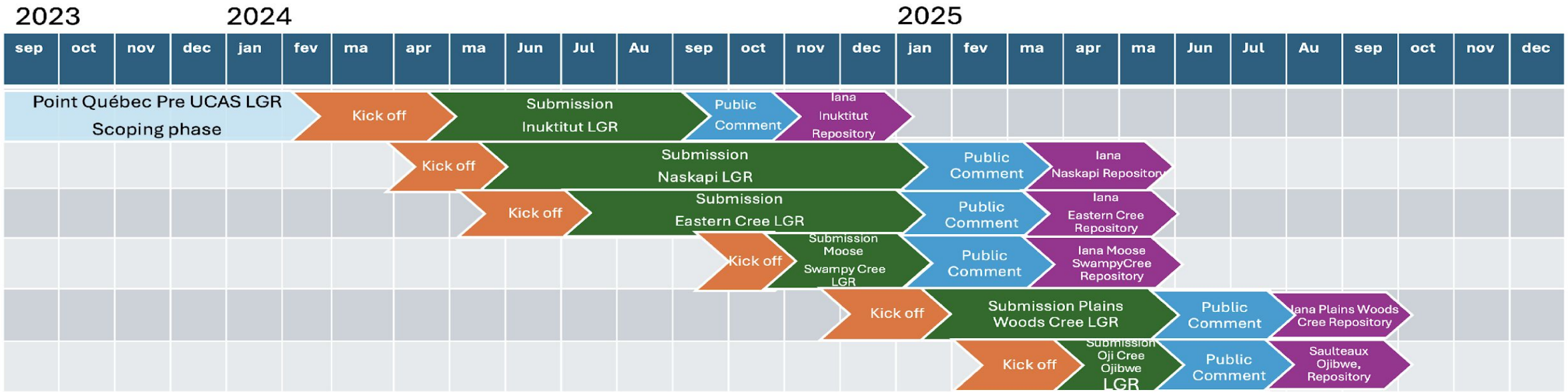
Louis Houle
Co-Chair, UCAS Language Reference LGR Panel

Background and Current Status

- Reference LGR for second-level domain names in official and recognized languages.
- Composition of the working group follows multistakeholder model.
  - Registry, Registry Operator, IT-engineering Specialists, Scholars, Linguistics and Typographic Experts, Community and governments.
- A multiple phases planned for languages using Unified Canadian Aboriginal Syllabics Script, as of September 2024, 4 LGR phases engaged:
  - Phase 1: Inuktitut (7 dialects)
  - Phase 2: Naskapi
  - Phase 3: Northern East, Southern East Cree
  - Phase 4: Moose Cree, Eastern Swampy Cree, Western Swampy Cree
- Status of Inuktitut language Reference LGR
  - Online meetings February – June 2024.
  - Published for public for public comment, closed 24 October 2024.
  - Final version after public comment available at https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en.

UCAS Reference LGR GP Timeline

UCAS - ICANN WG LGR Timeline



Phase 1 – Inuktitut
Phase 2 – Naskapi
Phase 3 – Southern Eastern Cree - Northern Eastern Cree
Phase 4 – Moose Cree - Eastern Swampy Cree - Western Swampy Cree
Phase 5 – Plains Cree – Woods Cree

Phase 6 – Oji-Cree – Ojibwe - Saulteaux
Phase 7 – Dakelh (Carrier) *Expected : fall 2025*
Phase 8 – Dene *Expected : fall 2025*

As of October 23, 2024

Moving toward Inclusive Internet

- The ICANN community promotes an Inclusive Internet for all languages in Internationalized Domain Name addresses.
- Browsers must be fully compatible with IDNs.
- Registries, registrars and Internet service providers (ISP) must upgrade their services to be IDN compatible and email address internationalized (EAI).
- Governments must promote inclusiveness by using local languages in their communications.
- In Québec Syllabics SLD deployment availability are planned for Q3-4 2025.

Introduction of IDN topics at ICANN 81 Meeting Week

IDN Related Sessions at ICANN 81 Meeting Week

- GDS: SubPro IRT IDN Sub-Track Work Session
  Monday November 11, 2024 13:15 - 14:30 GMT+03 (10:15 UTC)

- Progress on IDN Related Projects for the New gTLD Program: Next Round
  Wednesday November 13, 2024 15:00 - 16:00 GMT+03 (12:00 UTC)
  - IDN Implementation in the Next Round
  - RZ-LGR Current Version and Upcoming Updates
  - String Similarity Review
  - LGR Tool

Thank You

Q&A