Expedited Policy Development Process on Internationalized Domain Names (EPDP on IDNs)

Presentation #3 on the Phase 1 Initial Report

Satish Babu
Abdulkarim Oloyede
Hadia Elminiawi
Justine Chew

17 May 2023

ICANN
AT-LARGE

Agenda

- Recall: EPDP's remit is focused on **Variant Management policies**
- Substance (for today)
  - **Sec 4.4: Modifying String Similarity Review** to account for variants at TL
  - **Sec 4.6: String Contention**
  - ***Time permitting*, sec 4.3**:
    - Impact on Application Process & Fee-Related questions
    - Impact on Reserved Names & String ineligible for delegation

Resources

- Link to Initial Report: https://itp.cdn.icann.org/en/files/internationalized-domain-names-idn/phase-1-initial-report-internationalized-domain-names-expedited-policy-development-process-24-04-2023-en.pdf
- Additional informational slides
  - Understanding Variants: The Basics
  - Root Zone Label Generation Rule (RZ-LGR) and as the sole source to determine valid top-level domain labels, their variant labels, and disposition values of the variant labels (PR 1.1)
  - Not all scripts have variants
  - 4 Underlying Principles guiding this EPDP's PRs & IGs
  - Impact on Application Submission & Application Fees (PR 3.11, PR 3.12, PR 3.13 & PR 3.14)

RECAP: IDNs EPDP Phase 1 Initial Report Sections

4.1 RZ-LGR as the Sole Source

4.2 Same Entity Principle

4.3 Application Submission, Administrative Check, Initial Evaluation

4.4 String Similarity Review

4.5 Objection Processes

4.6 String Contention

4.7 Contractual Requirements

4.8 Delegation and Removal

4.9 Variant Label States

4.10 Charter Questions with No Preliminary Recommendations

RZ-LGR as Sole Source to determine Variant Label Set

- **PR 1.1**: The RZ-LGR will be the sole source to determine valid top-level domain labels, their variant labels, and disposition values of the variant labels.

A real example of RZ-LGR output for an Arabic label

Allocatable means available for delegation but must still be applied for delegation

Primary (label 1): ---------->
The label that is the source for calculating the variant label set and determining its variant labels that are allocatable or blocked in accordance with the RZ-LGR

Allocatable (labels 2,7,9,10,15,17,18,23): ------>
A valid variant label eligible to be a top-level domain and available for application, allocation, and eventual delegation

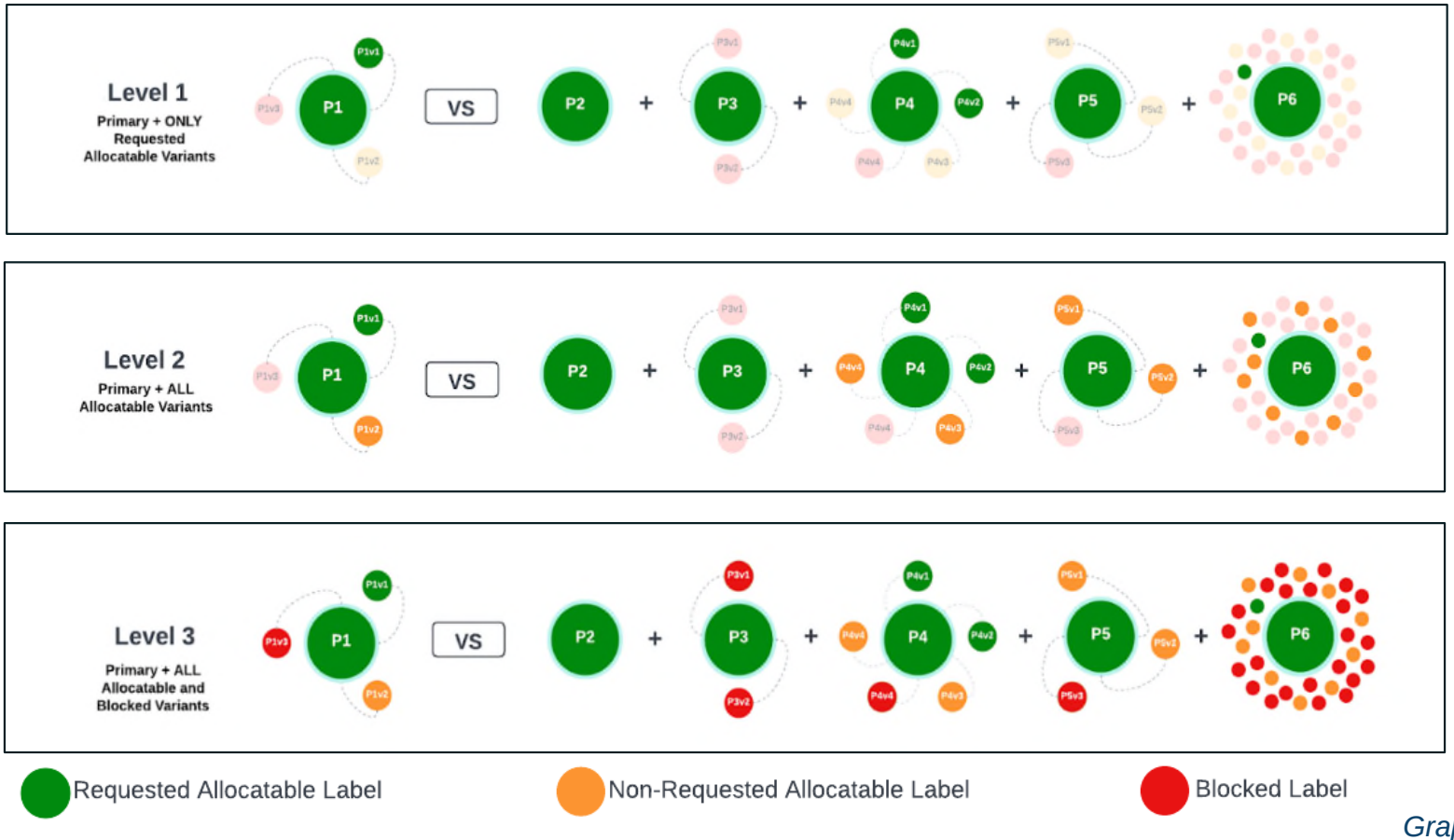
Blocked (labels 3-6,8,11-14,16,19-22, 24):->
A valid variant label not eligible for allocation or delegation as a top-level domain

Variant Label Set (labels 1-24) ---------->
The set of labels that is calculated by the RZ-LGR using the primary label, which consists of: primary label + allocatable variant label(s) + blocked variant label(s).

#	Type	U-label	A-label	Disposition	Code point sequence
1	original	شبكة	xn--ngbc5azd	valid	U+0634 U+0628 U+0643 U+0629
2	varlabel	شبكه	xn--ngbx0cq	allocatable	U+0634 U+0628 U+0643 U+0647
3	varlabel	شبكهٖ	xn--ngbx0c15a	blocked	U+0634 U+0628 U+0643 U+06BE
4	varlabel	شبكةٰ	xn--ngbx0c95a	blocked	U+0634 U+0628 U+0643 U+06C0
5	varlabel	شبكہ	xn--ngbx0cy6a	blocked	U+0634 U+0628 U+0643 U+06C1
6	varlabel	شبكۃ	xn--ngbx0c26a	blocked	U+0634 U+0628 U+0643 U+06C2
7	varlabel	شبكۃ	xn--ngbx0c66a	allocatable	U+0634 U+0628 U+0643 U+06C3
8	varlabel	شبكه	xn--ngbx0c31b	blocked	U+0634 U+0628 U+0643 U+06D5
9	varlabel	شبكة	xn--ngbc5az1b	allocatable	U+0634 U+0628 U+06A9 U+0629
10	varlabel	شبكه	xn--ngbx2d5u	allocatable	U+0634 U+0628 U+06A9 U+0647
11	varlabel	شبكهٖ	xn--ngbx66ayc	blocked	U+0634 U+0628 U+06A9 U+06BE
12	varlabel	شبكةٰ	xn--ngbx66a6c	blocked	U+0634 U+0628 U+06A9 U+06C0
13	varlabel	شبكہ	xn--ngbx66agd	blocked	U+0634 U+0628 U+06A9 U+06C1
14	varlabel	شبكۃ	xn--ngbx66akd	blocked	U+0634 U+0628 U+06A9 U+06C2
15	varlabel	شبكۃ	xn--ngbx66aod	allocatable	U+0634 U+0628 U+06A9 U+06C3
16	varlabel	شبكه	xn--ngbx66a0f	blocked	U+0634 U+0628 U+06A9 U+06D5
17	varlabel	شبكة	xn--ngbc5a31b	allocatable	U+0634 U+0628 U+06AA U+0629
18	varlabel	شبكه	xn--ngbx2d9u	allocatable	U+0634 U+0628 U+06AA U+0647
19	varlabel	شبكهٖ	xn--ngbx96asc	blocked	U+0634 U+0628 U+06AA U+06BE
20	varlabel	شبكةٰ	xn--ngbx96a0c	blocked	U+0634 U+0628 U+06AA U+06C0
21	varlabel	شبكہ	xn--ngbx96a4c	blocked	U+0634 U+0628 U+06AA U+06C1
22	varlabel	شبكۃ	xn--ngbx96a8c	blocked	U+0634 U+0628 U+06AA U+06C2
23	varlabel	شبكۃ	xn--ngbx96ahd	allocatable	U+0634 U+0628 U+06AA U+06C3
24	varlabel	شبكه	xn--ngbx96arf	blocked	U+0634 U+0628 U+06AA U+06D5

String Similarity Review: Modify to “Hybrid Model” (1/4)

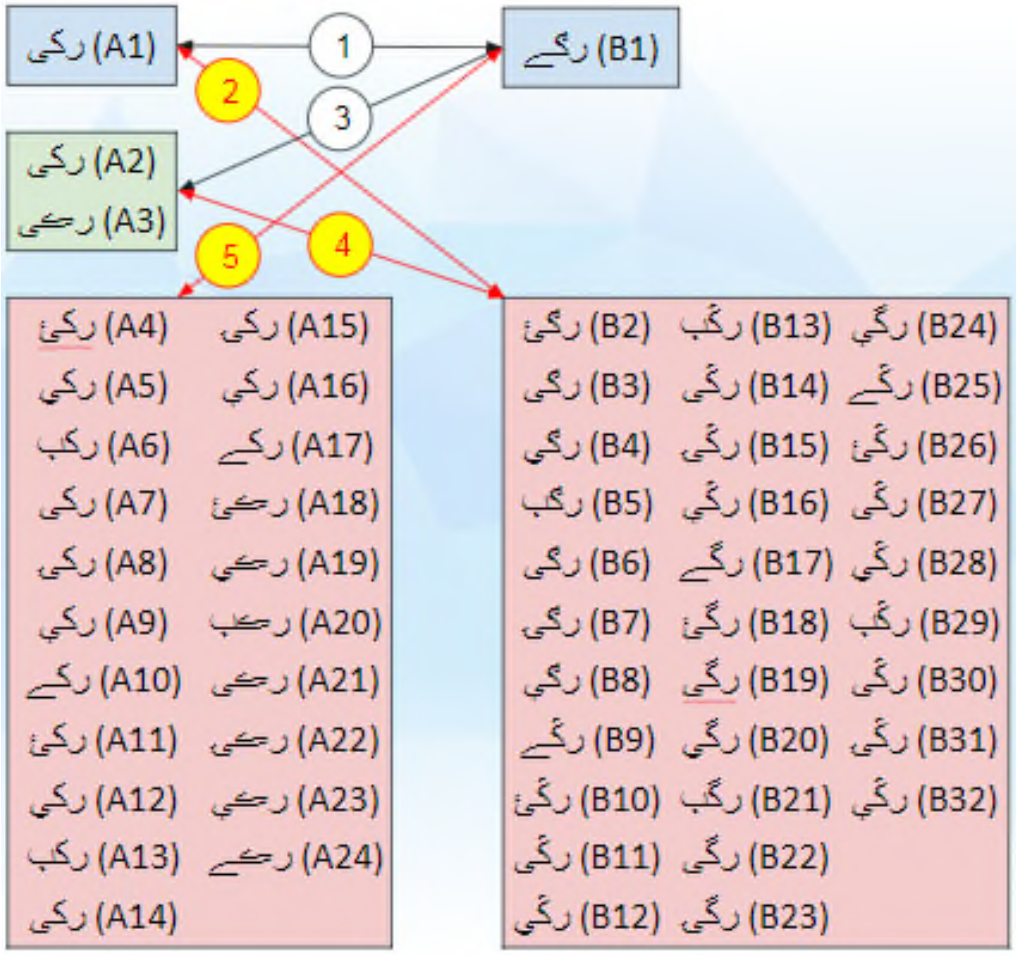
- String Similarity Review:
  - Visual test by String Similarity Review Panel (SSRP)
  - Identifies common-script, cross-script strings which are visually confusable
  - Takes place prior to objection process
  - What roles do variant labels play?



Graphic by Ariel Liang

String Similarity Review: Modify to “Hybrid Model” (2/4)

- **PR 4.1**: Modify 2012 String Similarity Review to Hybrid Model – compare all levels of strings against each other **except for** blocked against blocked



May find the following confusingly similar labels...

2 رکی (A1) & رگی (B3) & رگی (B6)

4 رکی (A2) & رگی (B3) & رگی (B6)

4 رے (A3) & رگی (B3) & رگی (B6)

5 رگے (B1) & رکے (A10) & رکے (A17) & رکے (A24)

Potential outcome...

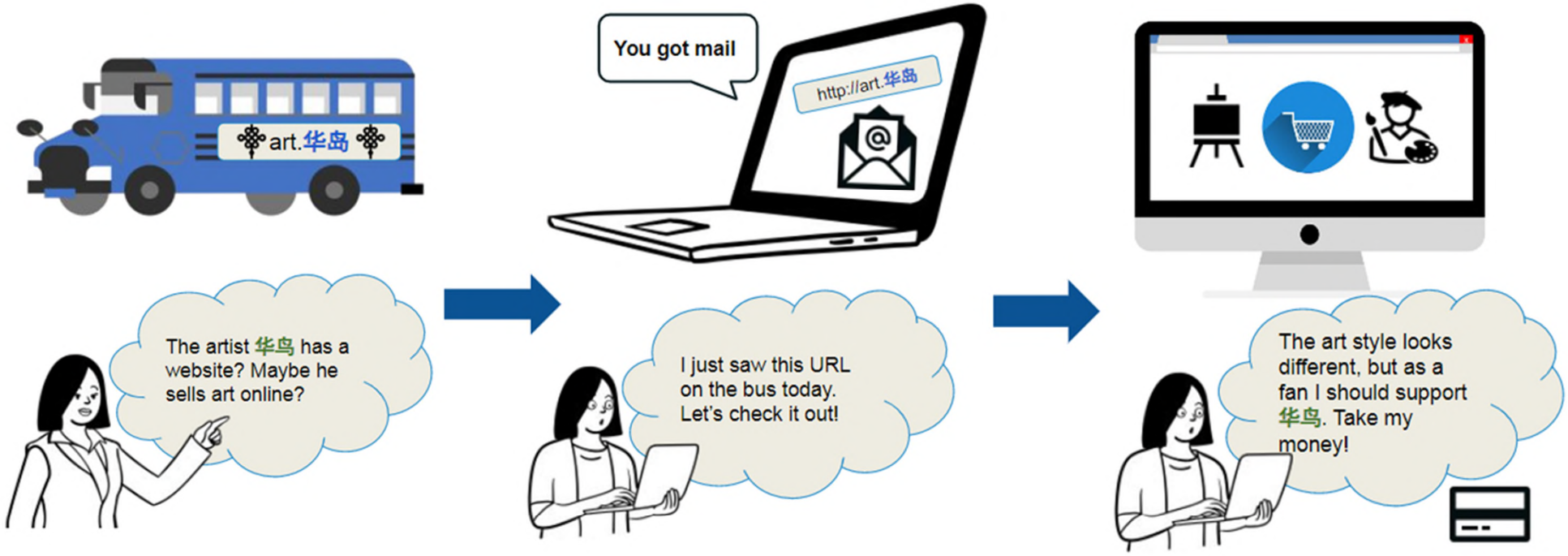
رکی (A1) & its variants A2-A24 AND رگے (B1) & its variants B2-B32 get processed in a contention set

If the Hybrid Model were not used...

رکی (A1) and رگے (B1) would have been both delegated with the misconnection risk. E.g., a user may mistake رکی (A1) as رگی (B3), a blocked variant of رگے (B1), but arrive at site controlled by a registrant different to رگے (B1).

String Similarity Review: Modify to “Hybrid Model” (3/4)

- Misconnection Risk & Potential Consequences
  - A user attempts to visit http://example.X, reading it as being the same as the http://example.Y that, for example, he or she saw in an advertisement. After clicking on http://example.Y, the user arrives at a site controlled by a registrant different to http://example.X.



- Misconnection **may** be more problematic than denial of service, **cause more harm to end-user beyond confusion and frustration**
- Arriving at the wrong site, even if legitimate**, can result in credential compromise and accidental exposure of information**
- If confusing similarity is maliciously leveraged, it **can be a DNS abuse vector. When confusion is at the TL, the possibility of DNS abuse is much greater** than that at the SL

Graphic by Ariel Liang

String Similarity Review: Modify to "Hybrid Model" (4/4)

- **PR 4.1**: Modify 2012 String Similarity Review to Hybrid Model – compare all levels of strings against each other **except for** blocked against blocked
  - Mitigate potential risks from denial of service and/or **misconnection**
  - Detects more combinations of visually confusable labels than
  - Avoids unnecessary complexity by not requiring blocked-blocked comparisons
  - **Presented to CPWG Call of 10 Oct 2022, with support received**

- **PR 4.2**: Allow SSRP to decide whether/what blocked variant labels to omit in SSR
  - Omission must be based on guidelines / criteria on basis of manifestly low level of confusability between scripts, additional research / study to identify such scripts

- **PR 4.3**: PR 4.2 guidelines/criteria must be developed during implementation

- **PR 4.4**: All labels from a variant label set must share same outcome of SSR
  - If applied-for variant label set & existing TLD exhibit confusing similarity → entire variant label set is ineligible to proceed
  - If applied-for variant label set & another applied for variant label set exhibit confusing similarity → both sets placed in a contention set. (PR 6.2)
  - NB. Anything not caught by SSR, there is String Confusion Objection as fallback

String Contention

- **Integrity of the Set:** The relationship between a primary label and its allocatable and blocked variant labels shall not be infringed upon as long as the primary label exists.

- Consequently:

  - **PR 6.1**: An applied-for primary gTLD string that is also a variant label of another applied-for primary gTLD string, as calculated by the RZ-LGR, must be placed in a contention set.

  - **PR 6.2**: The entire variant label set of an applied-for primary gTLD string (no matter whether it is an ASCII string or an IDN string) must be processed in the contention set

Time Permitting …

Let’s pick up from CPWG 3 May 2023 presentation

Application Process & Fee-Related PRs & IGs (1/3)

- EPDP Team considered the 2012 Round application & evaluation process flow – conclusion: not feasible (operationally & cost-wise) to have a “separate round” or separate application & evaluation process for variant labels
  - Too many of the existing processes – retained by SubPro – meant that we could not disregard them for variant labels
- Therefore:
  - **PR 3.2**: Future registry operator can only apply for allocatable variant label during application round
  - **PR 3.3**: Existing IDN gTLD registry operators can only apply allocatable variant labels during application round
    - With **PR 3.15**: One-time exception in the immediate next application round, existing IDN gTLD applications for allocatable variant labels to receive priority in processing order

Application Process & Fee-Related PRs & IGs (2/3)

- **Conservatism:** Adopt a more cautious approach in gTLD policy development as way to limit any potential security & stability risks associated with the variant label delegation.

- Led to measures to help ensure "safety & security" for end-users:
  - **PR 3.5**: Both future IDN gTLD and existing registry operators who want allocatable variant labels must explain why they seek those variant label
    - **IG 3.6**: Criteria for evaluating explanations (per PR 3.5) should be pre-identified and applied consistently by qualified evaluators
  - **PR 3.7**: Both future IDN gTLD and existing registry operators who want allocatable variant labels must demonstrate ability to manage primary and variant labels from technical and operational perspective
    - **IG 3.8**: Evaluation (per PR 3.7) should be closely tied to overall technical capability evaluation with criteria including Critical Functions with respect to SL registrations
    - **IG 3.9**: ICANN org may do research to help identify additional standards or test for technical and operational capability evaluation (per PR 3.7)

Application Process & Fee-Related PRs & IGs (3/3)

- **PR 3.4**: Future IDN gTLD primary and allocatable variants labels in one application
- **PR 3.10**: Fee structure for all future applications must be consistent with principle of cost recovery (SubPro)
- **PR 3.11, PR 3.12, PR 3.13 & PR 3.14** impact on application fee structure

Apply for	Next Round	A Future Round After Next Round
New Applicant		
Primary label only	Base Application Fee	Base Application Fee
Primary label + ≤ 4 variant labels	Base Application Fee	Base Application Fee
Primary label + > 4 variant labels	Base Application Fee + (May Incur) Additional Fees	Base Application Fee + (May Incur) Additional Fees
Existing Registry Operator from 2012 Round		
≤ 4 variant labels	Base Application Fee Waived	Discounted Base Application Fee
> 4 variant labels	Base Application Fee Waived + (May Incur) Additional Fees	Discounted Base Application Fee + (May Incur) Additional Fees
Future gTLD Registry Operator		
≤ 4 variant labels	Not Allowed	Discounted Base Application Fee
> 4 variant labels	Not Allowed	Discounted Base Application Fee + (May Incur) Additional Fees

Reserved Names & String Ineligible for Delegation

- **Reserved Names**
  - What: ICANN, ICANN bodies/groups, or related to ICANN functions and IANA
  - Egs: ALAC, ICANN, RIPE, GAC, CCNSO, GNSO, IAB, IETF, IANA, PTI etc
  - All the RNs, except of IDN “test” strings, are ASCII strings with only blocked variant labels
  - **PR 3.18**: Reserved Names list to not be expanded to include variant labels
  - **PR 3.19**: Variant labels of Reserved Names not allowed

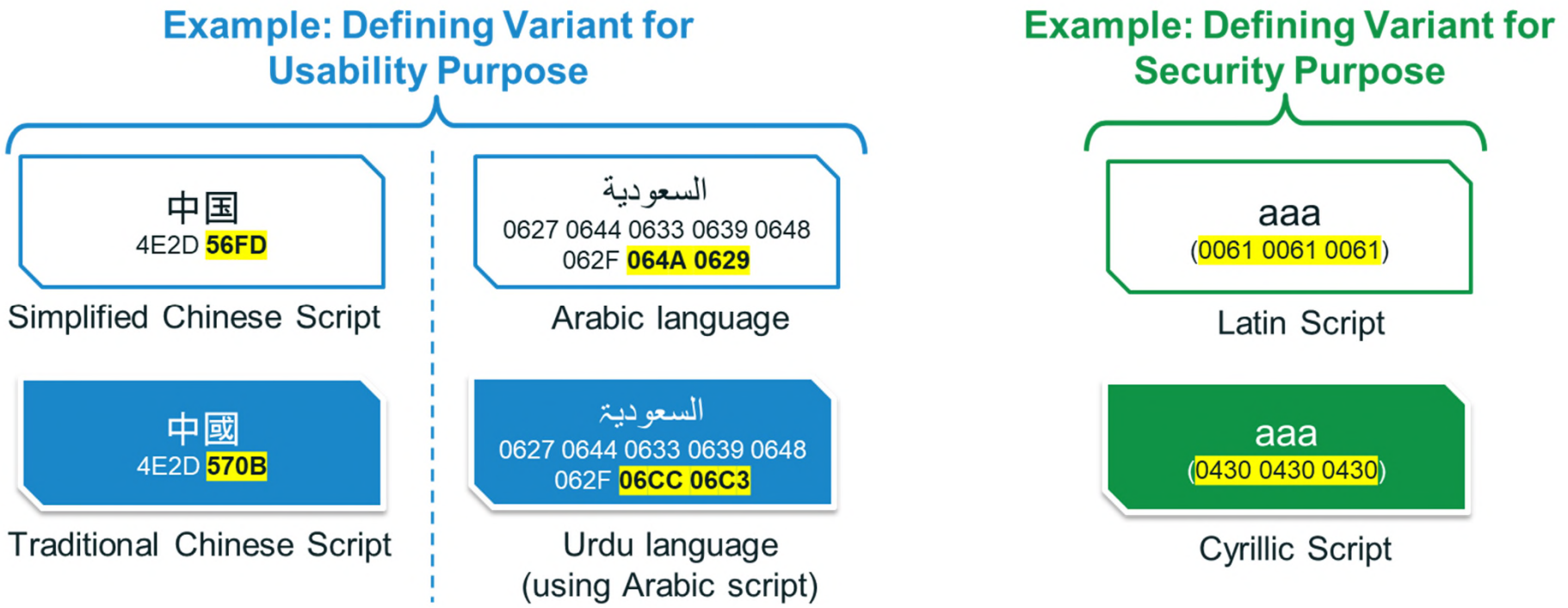
- **Strings ineligible for delegation**
  - What: special protections at TL & SL for names, acronyms of IGOs, INGOs with protections under treaties and statutes across multiple jurisdictions
  - Egs: Red Cross/Red Crescent Movement (RCRC); Int Olympic Comm (IOC)
  - **PR 3.20**: List of Strings Ineligible for Delegation to not be expanded to include variant labels
  - **PR 3.21**: Only the protected orgs on list of Strings Ineligible for Delegation can apply variant labels of their protected strings; but only if they also apply for or have the primary

End

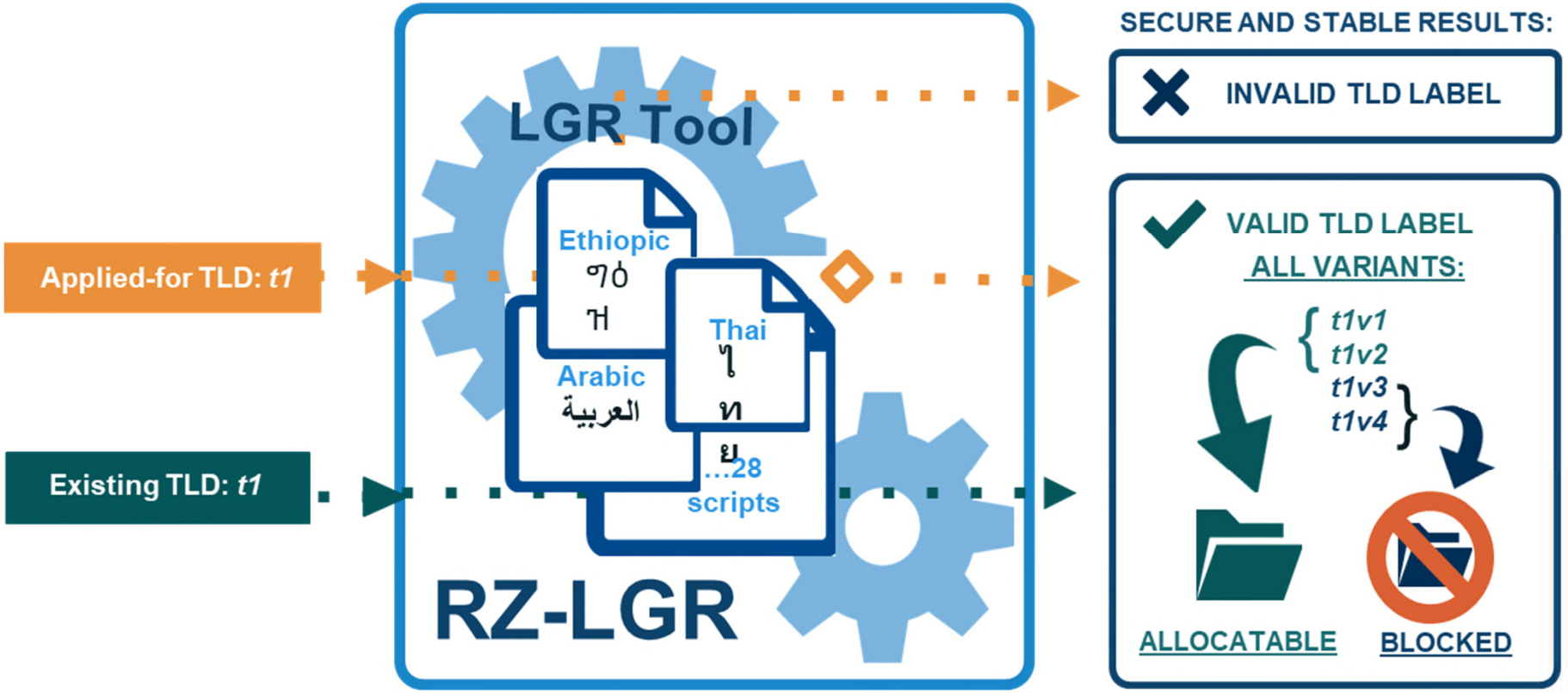
Thank you for your input.

Understanding Variants: The Basics

- **Variant Labels** are considered **'the same**' by respective script community



Root Zone Label Generation Rules (RZ-LGR)



Total number of script communities (Generation Panels): 17

Total number of participant across script communities: 270+

Total number of languages represented: 386+

Total number of population represented: 5 billions

Total number of hours worked (estimated): 10,000+ hours

Total number of LGRs developed: 25

Not all scripts have variants

4 Underlying Principles

- **RZ-LGR as the Sole Source:** The RZ-LGR will be the sole source to determine valid top-level domain labels, their variant labels, and disposition values of the variant labels. (Subject of PR 1.1)

- **Same Entity:** At the top-level of the DNS, the same registry operator must manage the approved labels from the variant label set of a primary gTLD from the application, legal, and operational standpoints. (Subject of PR 2.1)

- **Integrity of the Set:** The relationship between a primary label and its allocatable and blocked variant labels shall not be infringed upon as long as the primary label exists.

- **Conservatism:** Adopt a more cautious approach in the gTLD policy development as a way to limit any potential security and stability risks associated with the variant label delegation.

See: Section 3: Glossary