ICANN

No. 49
23-27 March 2014
Singapore

IDN Variant TLD Program Update

Wednesday, 26 March 2014

Agenda

- Program Update (15 minutes)

- Maximal Starting Repertoire (20 minutes)

- Representing Label Generation Rulesets in XML (15 minutes)

- Community updates (20 minutes)

- Q&A (20 minutes)

IDN Variant Program: A Brief Overview

Phase 1: 2011

- Script Case Studies Conducted: Arabic, Chinese, Cyrillic, Devanagari, Greek, Latin

IDN Variant Program: A Brief Overview

Phase 1: 2011

o Script Case Studies Conducted: Arabic, Chinese, Cyrillic, Devanagari, Greek, Latin

Phase 2: 2011 – 2012

o Integrated Issues Report development and publication

IDN Variant Program: A Brief Overview

Phase 1: 2011 ✓

- Script Case Studies Conducted: Arabic, Chinese, Cyrillic, Devanagari, Greek, Latin

Phase 2: 2011 – 2012 ✓

- Integrated Issues Report development and publication

Phase 3: 2012 – 2013 ✓

- Creation of Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA labels (LGR Procedure)

- Development of "Study on Examining the User Experience Implications of Active Variant TLDs" and

- Specification for representing Label Generation Rulesets (LGR) in XML

IDN Variant Program: A Brief Overview

Phase 1: 2011
- Script Case Studies Conducted: Arabic, Chinese, Cyrillic, Devanagari, Greek, Latin

Phase 2: 2011 – 2012
- Integrated Issues Report development and publication

Phase 3: 2012 – 2013
- Creation of Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA labels (LGR Procedure)
- Development of "Study on Examining the User Experience Implications of Active Variant TLDs" and
- Specification for representing Label Generation Rulesets (LGR) in XML

Phase 4: 2013 – 2015 (In Progress)
- Implementation of LGR procedure
- Processes development for incorporating the LGR
- Ongoing work on XML specification

IDN Variant Program: A Brief Overview

Phase 1: 2011 ✓

- Script Case Studies Conducted: Arabic, Chinese, Cyrillic, Devanagari, Greek, Latin

Phase 2: 2011 – 2012 ✓

- Integrated Issues Report development and publication

Phase 3: 2012 – 2013 ✓

- Creation of Procedure to develop and maintain the Label Generation Rules for the Root Zone in Respect of IDNA labels (LGR Procedure)

- Development of "Study on Examining the User Experience Implications of Active Variant TLDs" and XML specification for representing Label Generation Rules

Phase 4: 2013 – 2015 (In Progress)

- Implementation of LGR procedure

- Processes development for incorporating the LGR

- Ongoing work on XML specification

LGR Procedure Overview

Generation Panels

- Generate proposals for script specific LGRs, based on community expertise and requirements
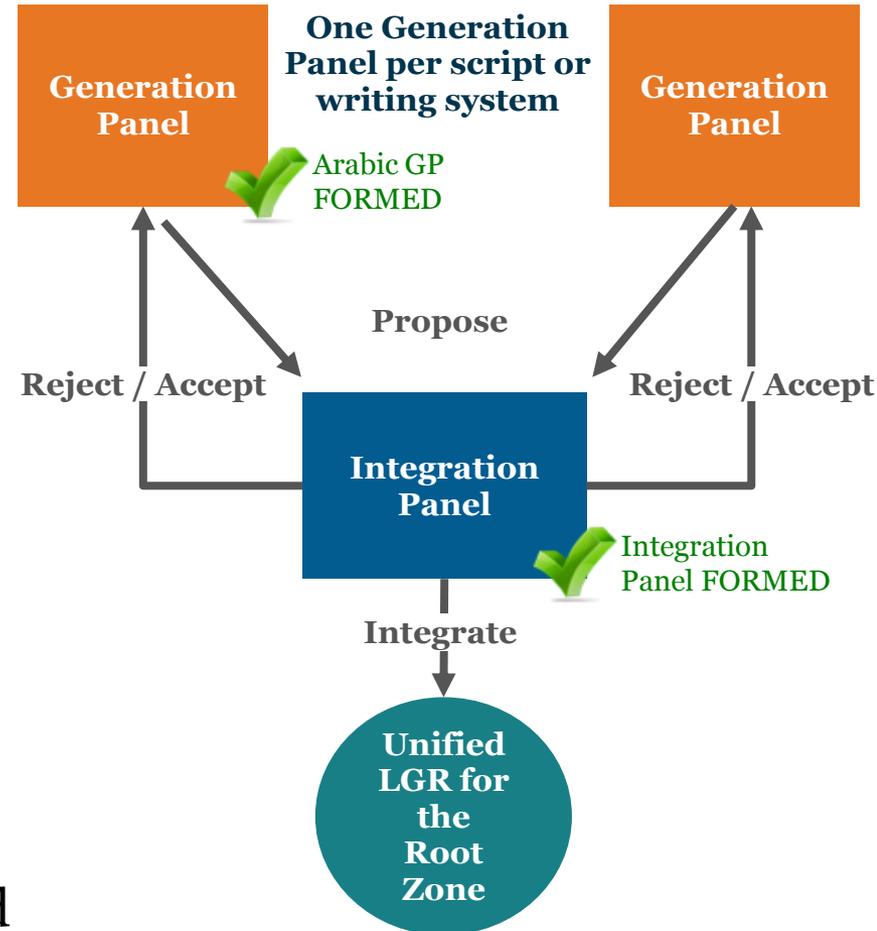
Integration Panel

- Integrates them into common Root Zone LGR while minimizing the risk to Root Zone as shared resource

Label Generation Rules (LGR)

- Which labels are permissible

- Which variant labels exist

- Which variant labels may be allocated

TO BE FORMED BY SCRIPT COMMUNITIES

Generation Panel

One Generation Panel per script or writing system

Arabic GP FORMED

Generation Panel

Propose

Reject / Accept

Reject / Accept

Integration Panel

Integration Panel FORMED

Integrate

Unified LGR for the Root Zone

* URLs available on Slide 10 (Resources) and last slide

Implementation of LGR Procedure

(Assumes no changes needed following Public Comment)

⚠️ Timeline dependent on Community Work

Community Work

Completed

→ Submission of proposal to form Generation Panel for Arabic Script to ICANN

→ Generation Panel for Arabic script seated

Not Started

→ Submission of additional proposals to form Generation Panels

→ Additional Generation Panels seated

ICANN Work

Completed

→ Publication of:
- Call for Root Zone LGR Generation Panels
- Call to form the Integration Panel/Advisors Pool

→ Integration Panel 1st F2F meeting & startup tasks

→ LGR Integration Panel seated

In Progress

→ Publication of:
- Maximal Starting Repertoire (MSR) V01 for Public Comment: Close 21 May 2013
- MSR V01

Submission of proposed script LGRs

Finalizing Label Generation Rules

Not Started

→ Public Comment: X LGR

→ Draft Integrated LGR

→ Integrated LGR V01

→ ICANN Board Review of Integrated LGR V01

Maximal Starting Repertoire (MSR): According to the Procedure, the integration panel is tasked with establishing the maximal set of code points and default whole label variant evaluation rules for the root zone, which serve as a starting point for generation panels. The maximal set of code points for the root zone is itself a subset of Unicode created via an application of IDNA2008 and the principles in IABCP.

Label Generation Rules (LGR): Every zone on the Internet has, either implicitly or explicitly, a set of rules governing the labels allowed in that zone, called label generation rules, or LGR in short. The Procedure document provides a mechanism to generate some of those rules: the ones necessary for long-term operation of both IDNs and IDN variants in the root zone.

Generation Panels Status

- Seated: Arabic

- Being Formed: Chinese, Japanese, Korean, Neo-Brahmi

- Expression of Interest: Cyrillic, Latin

LGR Procedure Depends on Community Work

- Without a community-based Label Generation Panel for a given script, there is no mechanism for determining IDN variant labels for that script

- Generation Panels need to represent the community and be comprised of experts on matters specific to a particular script

- **Generation Panels and LGR proposals are REQUIRED for IDN variants to be considered for delegation**

Arabic
Bengali
Chinese
Cyrillic
Devanagari
Georgian
Greek
Gujarati
Gurmukhi
Hebrew
Japanese
Korean
Latin
Sinhala
Tamil
Telugu
Thai

Get Involved

- Visit the LGR public workspace and the IDN Variant TLDs page for information on how to:

  - Form a generation panel

  - Volunteer to join a generation panel

  - Take part in public review of the MSR, LGR proposals, integrated LGR, etc.

  - Disseminate information to interested communities

ICANN Support for LGR Generation Panels

- Provides necessary documentation for getting started

- Provides a point of contact for groups wishing to form a panel or individuals wishing to join one

- Provides support to reach out to community members via ICANN channels

- Maintains the project website and mailing lists

- Provides a pool of advisors to help panels supplement required expertise

Resources

- Announcement: Call for Generation Panels to Develop Root Zone Label Generation Rules

  - http://www.icann.org/en/news/announcements/announcement-11jul13-en.htm

- Relevant documents:

  o Setting up and Running a Generation Panel

    - https://community.icann.org/download/attachments/43989034/SettingupandRunningaGenerationPanel.pdf

  o Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels

    - http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf

  o IDN Variant TLD Program Timeline/Reports/Presentations

    - http://www.icann.org/en/resources/idn/variant-tlds

Resources (continued)

- Root Zone LGR Project Workspace

    - https://community.icann.org/display/croscomlgrprocedure/Project+Status+Updates

- Mailing lists:

    o LGR@icann.org: Discuss matters related to LGR work and submissions of LGR proposal. Contact Integration and Generation Panel members

    o idntlds@icann.org: Contact ICANN to submit Generation Panel statement of interests, work reports, updates, etc.

    o vip@icann.org: Discuss issues related to the IDN Variant TLDs Program

        - Subscribe here: https://mm.icann.org/mailman/listinfo/vip

Maximal Starting Repertoire (MSR)

Context for the MSR

- MSR (Maximal Starting Repertoire) and WLE (Whole Label Evaluation) are part of the Root Zone LGR (Label Generation Rules) development

  - <u>Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels</u>

- The LGR project will result in a set of rules:

  - Define permissible labels for the root

  - Can be mechanically applied (automated)

  - Expressed as a set of parallel and consistent per-script rules

  - The Integrated LGR applies to all scripts covered

Maximal Starting Repertoire

- One of the initial tasks of the Integration Panel

- Outer limit of code points allowed in the root zone

- Subset of IDNA 2008 PVALID code points
  - No digits / punctuation / context dependent / unstable
  - No historic / obsolete/ limited use

- Some code points may not be part of the final LGR
  - Based on further evaluation by Generation Panel
    - Some may ultimately be not acceptable
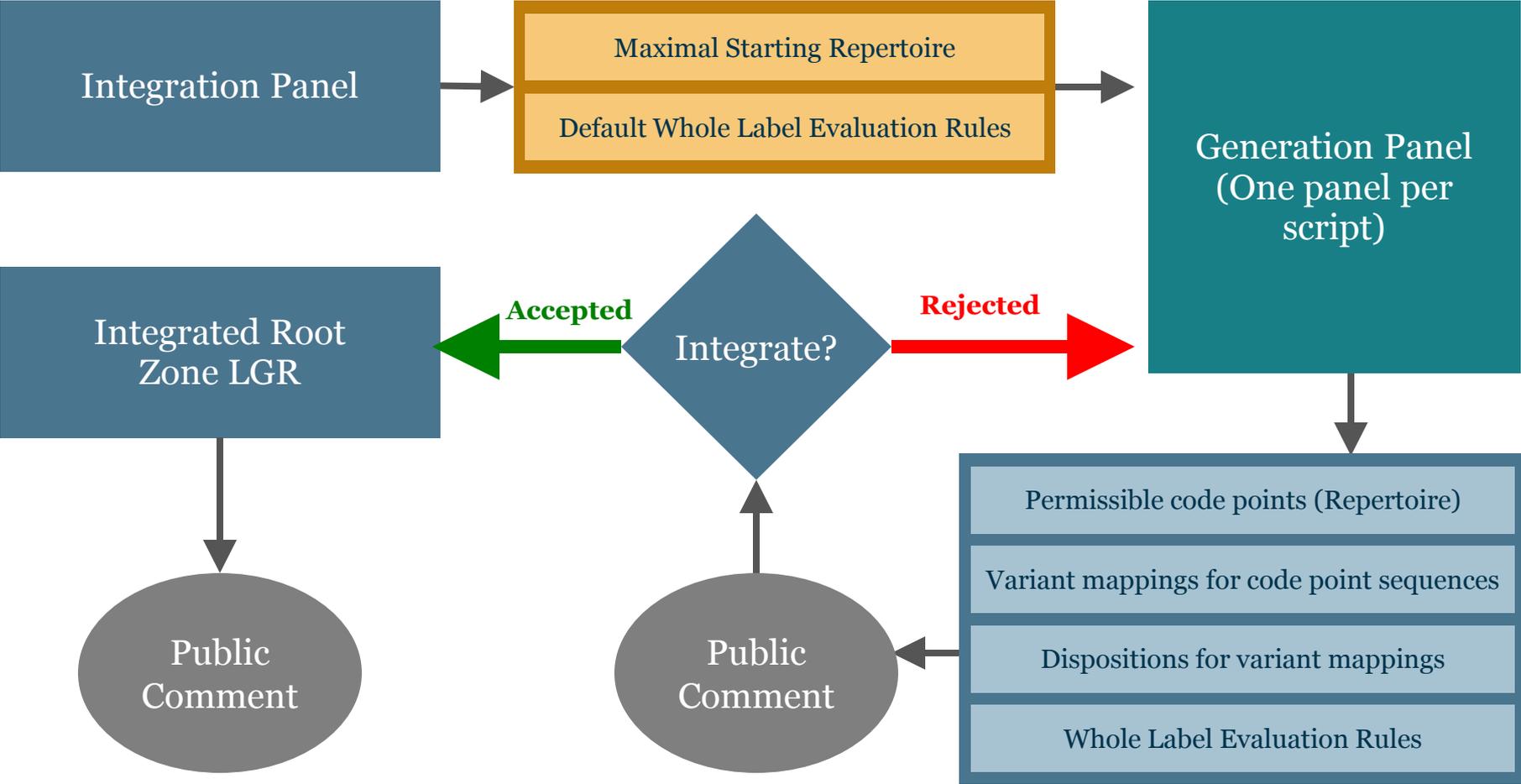   - Some may only be allowed in specific sequences or conditions

MSR-1 Status

- MSR-1: first Version of MSR

- Released for Public Comment 3 March 2014
  - http://forum.icann.org/lists/comments-msr-03mar14/

- Focuses on scripts for which IDN TLDs have been applied for
  - Also includes some closely related scripts

- Can be adjusted based on public comment

- To be followed later by MSR-2 when more scripts are added

After MSR-1 Is Finalized

- Following the LGR Procedure:

  o Generation Panels select code points from MSR to include in their script LGRs' repertoire

  o Provide justification for inclusion

  o Add other elements of LGR

- Ongoing dialog between GP and IP before submission of LGR proposal

- IP will review and accept/reject script LGRs for integration

Flow for a Script-based LGR

MSR-1 Content In Numbers

- 22 scripts
  - Cyrillic, Georgian, Greek, Latin
  - Arabic, Hebrew
  - Han Ideographs, Hangul syllables, Hiragana, and Katakana
  - Bengali, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Sinhala, Tamil, Telugu
  - Thai, Lao
- 'Common' and 'Inherited' (shared)
- 32,783 code points
  - 11,172 Hangul syllables and 19,849 Han ideographs

Whole Label Evaluation (WLE)

- MSR-1 contains a single default WLE rule

  - Prevents leading combining marks

- WLE are used to prevent ill-formed labels

- WLE Rules can be a method to manage variants

What About Variants?

- Potential variant examples[*]:
  - U+4C81: 鲥 U+9CDA: 鲥
  - Strasse, Straße
- Identifying code point variants and assigning dispositions is a task for **Generation Panels**
  - Not addressed in MSR
- Integration Panel will look for:
  - Justification for variants
  - Minimizing use of allocatable variants
  - Preferred use of blocked variants to prevent issues

* These are conceptual examples, not suggestions

Sample MSR Annotated Code Point Table

Archaic letters

0370	Ⱶ	GREEK CAPITAL LETTER HETA
0371	ⱶ	GREEK SMALL LETTER HETA
		• obsolete
0372	Ⲧ	GREEK CAPITAL LETTER ARCHAIC SAMPI
0373	ⲧ	GREEK SMALL LETTER ARCHAIC SAMPI
		• obsolete

Numeral signs

0374	ʹ	GREEK NUMERAL SIGN
0375	͵	GREEK LOWER NUMERAL SIGN
		• ContextO
		• obsolete

Archaic letters

0376	Ⱶ	GREEK CAPITAL LETTER PAMPHYLIAN DIGAMMA
0377	ⱶ	GREEK SMALL LETTER PAMPHYLIAN DIGAMMA

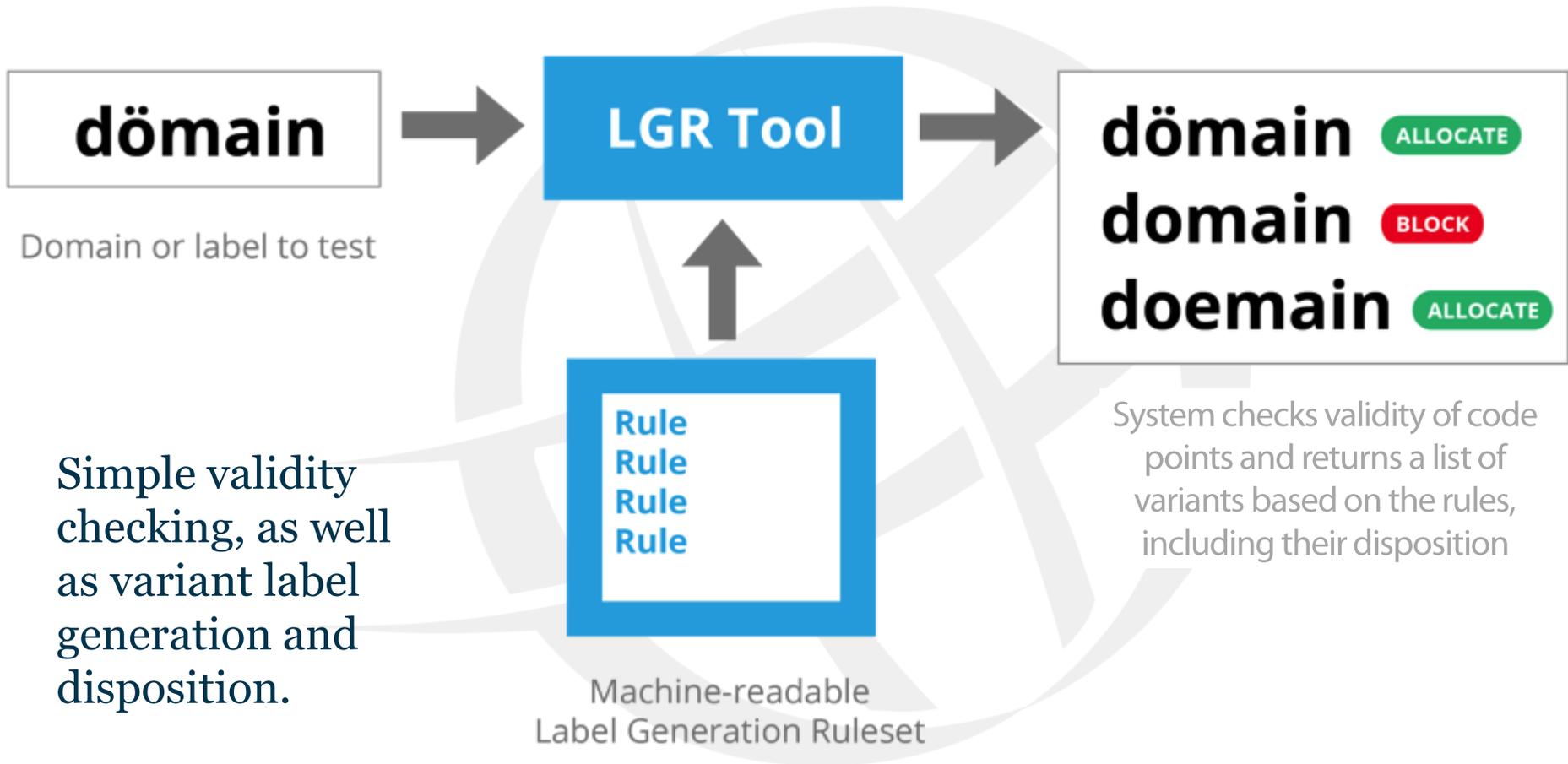|	
☐	Not PVALID
🟨	In MSR-1
🟪	Excluded from MSR-1

Representing Label Generation Rulesets (LGR) in XML

Features of XML-LGR Format

- Ability to express more of a registry's IDN policies beyond simple code point lists

- Provides comprehensive way to describe registry policies relating to:
  - ○ Permitted code points
  - ○ Definition and disposition of variants
  - ○ Rules for evaluation of whole label validity

- Universal format that can be implemented in a single fashion across multiple domains and policies
  - ○ XML format allows LGR to be machine processed

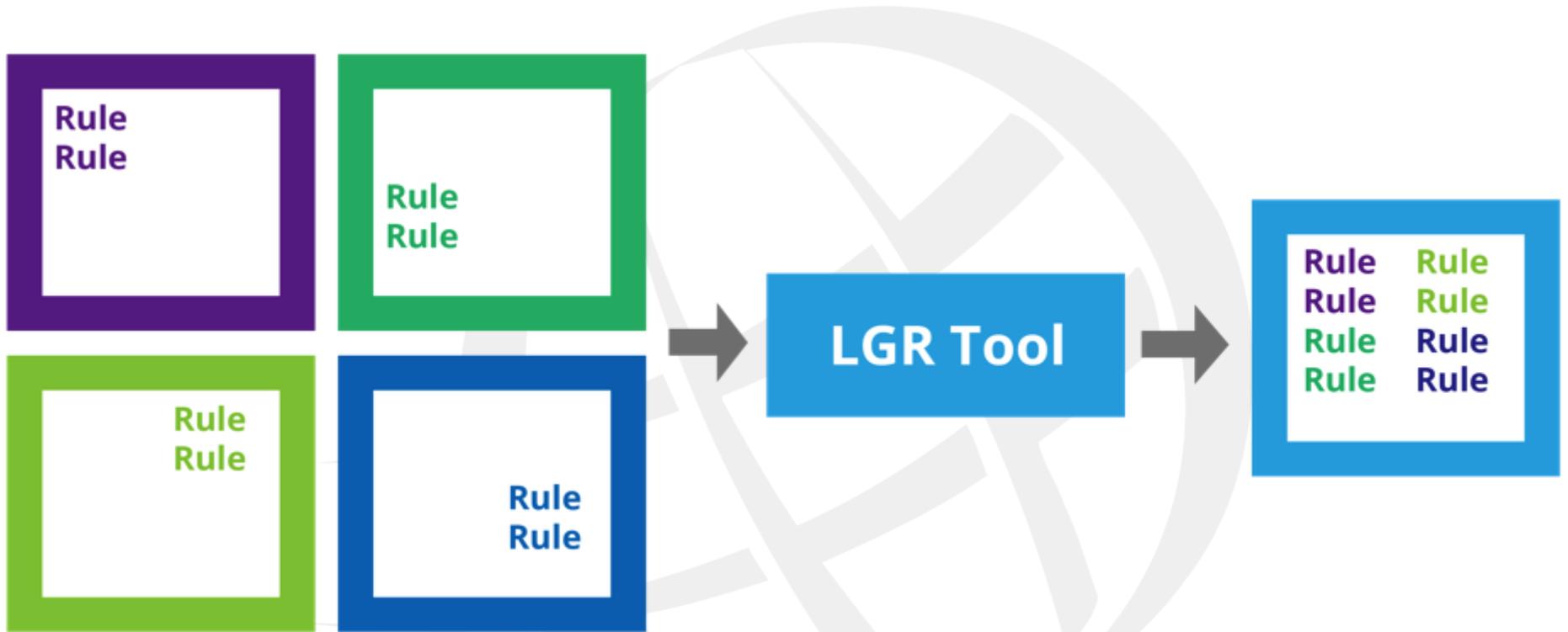- Good basis for clear and consistent rulesets for the root zone

What Does XML-LGR Enable?

domain
Domain or label to test

LGR Tool

domain ALLOCATE
domain BLOCK
doemain ALLOCATE

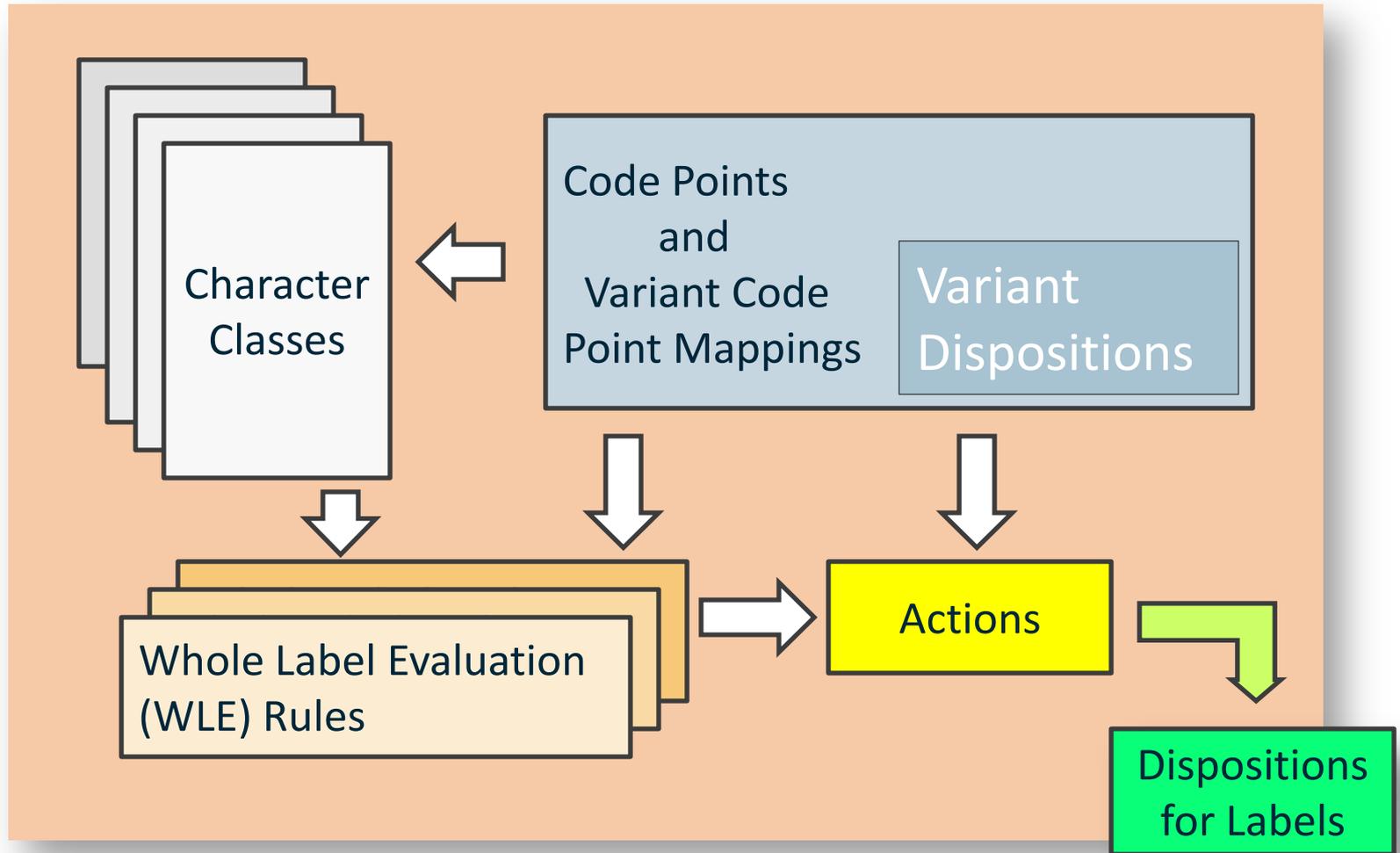System checks validity of code points and returns a list of variants based on the rules, including their disposition

Rule
Rule
Rule
Rule

Machine-readable Label Generation Ruleset

Simple validity checking, as well as variant label generation and disposition.

What Does It Enable?

Compare or merge tables from different sources

What does an LGR contain?

The Next Steps

- Finalize the specification and move tool(s) into production state

- Currently under review by Integration Panel

- Community review desirable:

  - http://tools.ietf.org/html/draft-davies-idntables

  - Send feedback or discuss on: public mailing list vip@icann.org

- Use the specification as the basis for Root LGR work

- Encourage existing implementers to migrate IDN table usage to LGR format

Creating LGRs or Converting to XML Format

- Generally a straightforward process

- The internet draft contains detailed examples

  - Including how to convert RFC 3743-style IDN tables

- MSR-1 can be used as a template for simple LGR with no variants

- Use available tools to check that XML is valid

  - Xmllint

  - RelaxNG <compact> schema for validation

- XML Format Tutorial in afternoon session

Community Work

Arabic Generation Panel

ICANN Singapore

Want To Know More? Join us for the LGR Workshop

- Title: IDN Root Zone LGR Workshop: Integration & Generation Panels

- Date/Time: Wednesday, 26 March 2014 — 13:00–17.00 SGT

- Room: Hullet

- Agenda:

  ○ Session 1: Community's role in creating the LGR (13.15–15.00)

    ▪ Guidance on how to form a Generation Panel and collaborate with the Integration Panel

    ▪ Community work to establish GPs

  ○ Session 2: Maximal Starting Repertoire and Whole Label Evaluation Rules (15.15–15.50)

  ○ Session 3: Training on XML format to represent the LGR (15.50–16.50)

ICANN IDN Team
Thank You

USEFUL LINKS:

- The LGR Procedure: http://www.icann.org/en/resources/idn/variant-tlds/lgr-procedure-20mar13-en.pdf

- MSR Version 1 Available for Public Comment: http://www.icann.org/en/news/public-comment/msr-03mar14-en.htm

- V06 Internet Draft for LGR Rules Toolset Project Published: http://tools.ietf.org/html/draft-davies-idntables

- Generation Panel for Arabic Script Root Zone Label Generation Rules Seated: http://www.icann.org/en/news/announcements/announcement-2-14feb14-en.htm

- Selection of Integration Panel for the IDN Root Zone Label Generation Rules: http://www.icann.org/en/news/announcements/announcement-06sep13-en.htm

- Call for Generation Panels to Develop Root Zone Label Generation Rules: http://www.icann.org/en/news/announcements/announcement-11jul13-en.htm

- Setting up and running a Generation Panel: https://community.icann.org/display/croscomlgrprocedure/Generation+Panels

- Community Wiki LGR Project website: https://community.icann.org/display/croscomlgrprocedure/Root+Zone+LGR+Project

ICANN Singapore