Internationalized Domain Names Expedited Policy Development Process

E3 - String Similarity Org Input
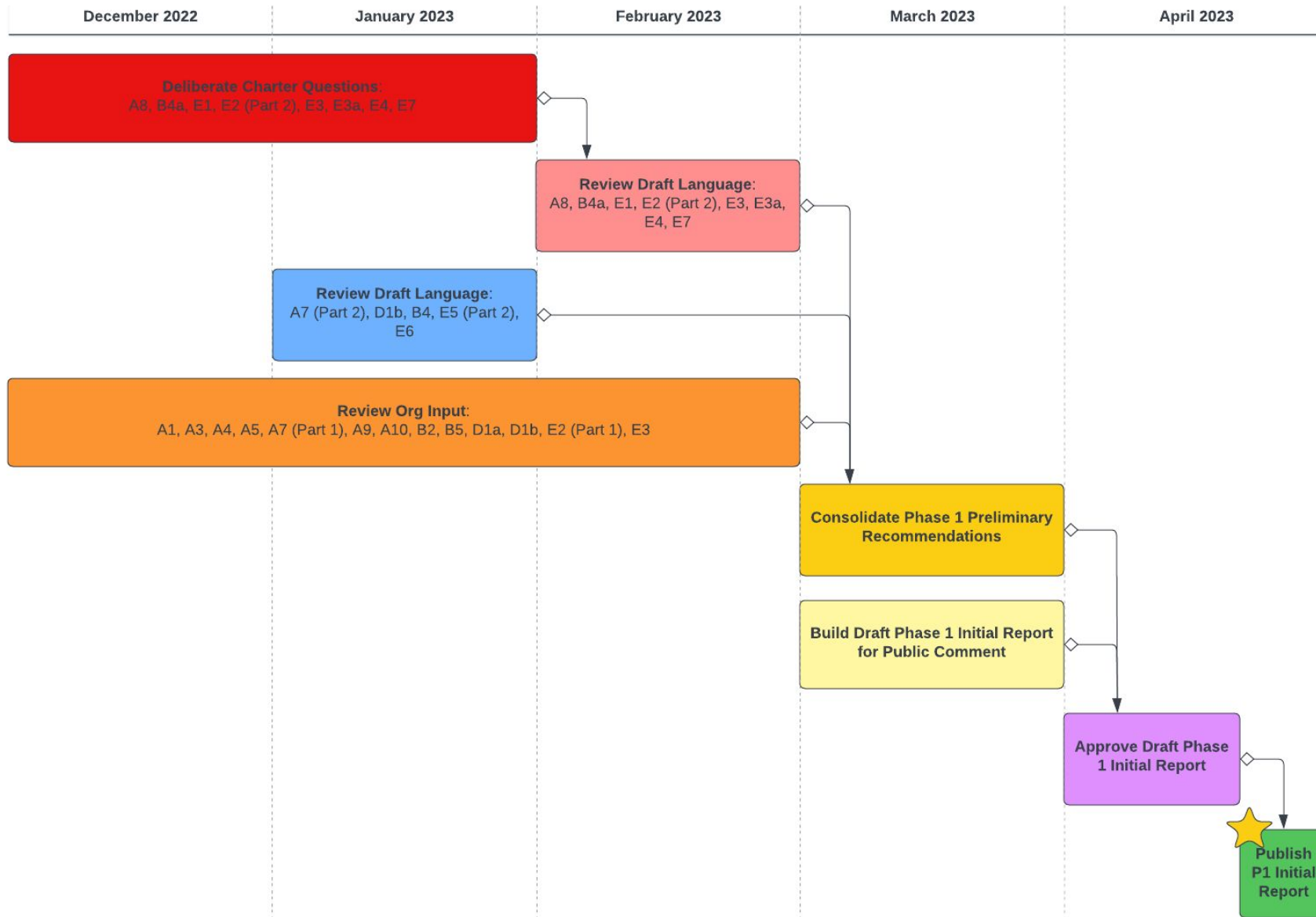
IDN-EPDP Team Meeting #61 | 8 December 2022

ICANN

Agenda

1. Roll Call and SOI Updates (2 mins)

2. Welcome and Chair Updates (5 mins)

3. Review of critical path to Phase 1 Initial Report (15 mins)

4. ICANN org input on the String Similarity Review Hybrid Model (65 mins)

5. AOB (3 mins)

Review of critical path to Phase 1 Initial Report

Critical Path to Phase 1 Initial Report

Review Org Input on String Similarity Review Hybrid Model

Summary

- **Focus:** Analyze the potential operational impact of the proposed hybrid model on the new gTLD process

- **Overview:** ICANN org conducted an analysis to determine the potential number of comparisons that would need to be performed in the String Similarity Review, of the models the EPDP Team is considering, ie Levels 1-3 and Hybrid

- **Method:**

  - Randomly selected 20 gTLD strings from the 2012 round as example applied-for strings (skewed toward IDNs)

  - For simplicity, assumed only the primary string is being applied for, not any of its allocatable variant(s)

  - Calculated allocatable and blocked variant labels of the strings using RZ-LGR-5, with mixed script labels removed

	Compare <u>Source Label</u>	Against <u>Target Label</u>
Level 1	primary	primary
Level 2	primary + all allocatable	primary + all allocatable
Hybrid	primary + all allocatable	primary + all allocatable + all blocked
	all blocked	primary + all allocatable
Level 3	primary + all allocatable + all blocked	primary + all allocatable + all blocked

target label = a label that the source label needs to compare with in the string similarity review. For example, in Level 1, one source primary label will be compared against 19 target primary labels, as there are a total number of 20 gTLD primary labels in the pool.

Calculate the Number of Comparisons

#	Script	TLD	Source Label			Target Label			Number of Comparison			
			number of primary	number of allocatable	number of blocked	nubmer of remaining primary	number of remaining allocatable	number of remaining blocked	L1 =DxG	L2 =(D+E)x (G+H)	Hybrid =(D+E)x (G+H+I)+Fx (G+H)	L3 =(D+E+F)x (G+H+I)
1	Chinese	网站	1	1	-	19	6	509	19	50	1,068	1,068
2	Chinese	游戏	1	1	10	18	5	499	18	46	1,274	6,264
3	Chinese	网址	1	1	2	17	4	497	17	42	1,078	2,072
4	Chinese	商店	1	-	3	16	4	494	16	20	574	2,056
5	Chinese	购物	1	1	-	15	3	494	15	36	1,024	1,024
6	Arabic	همراه	1	1	268	14	2	226	14	32	4,772	65,340
7	Arabic	ارامكو	1	2	147	13	-	79	13	39	2,187	13,800
8	Arabic	بازار	1	-	24	12	-	55	12	12	355	1,675
9	Japanese	グーグル	1	-	4	11	-	51	11	11	106	310
10	Japanese	家電	1	-	3	10	-	48	10	10	88	232
11	Cyrillic	москва	1	-	-	9	-	48	9	9	57	57
12	Cyrillic	орг	1	-	2	8	-	46	8	8	70	162
13	Devanagari	नेट	1	-	1	7	-	45	7	7	59	104
14	ASCII	etisalat	1	-	23	6	-	22	6	6	166	672
15	ASCII	next	1	-	3	5	-	19	5	5	39	96
16	ASCII	stcgroup	1	-	7	4	-	12	4	4	44	128
17	ASCII	smart	1	-	1	3	-	11	3	3	17	28
18	ASCII	rexroth	1	-	1	2	-	10	2	2	14	24
19	ASCII	ladbrokes	1	-	3	1	-	7	1	1	11	32
20	ASCII	land	1	-	7	0	-	-	-	-	-	-
	Total		20	7	509				190	343	13,003	95,144

Results

- **Number of Comparisons:**

  - **Level 1**: 190 comparisons

  - **Level 2**: 343 comparisons

  - **Hybrid**: 13,003 comparisons

  - **Level 3**: 95,144 comparisons

- **Caveats:**

  - The number of comparisons in this calculation represents <u>theoretical limits</u>

  - In practice, the actual number of comparisons may be fewer, depending on the visual similarity among strings (e.g. the String Similarity Review panel may decide it is not necessary to compare an Arabic string with a Chinese string, etc.)

- **Takeaways:**

  - The mitigation of confusion risks will likely be enhanced from Level 1 to Level 3, but the cost of operating the String Similarity Review will likely increase from Level 1 to Level 3

  - The added costs will likely be passed onto applicants, given the cost recovery principle

  - Hybrid model is a compromise between Level 2 and Level 3 in reducing computational complexity involving blocked variants

String Similarity Process & Hybrid Model

Application / Admin Check

1. applicant begins application process
2. Applicant elects whether to designate application as community-based
3. Applicant submits application into system
4. ICANN publishes list of all complete applications

Initial Evaluation (IE) String Review

5. **ICANN runs algorithm** for all applied-for gTLDs against all other applied-for gTLDs
   - **Step 1. Identical strings** added to contention sets
   - **Step 2. Strings with variant relations according to RZ-LGR** added to contention sets
6. **String Similarity Panel performs analysis**, using algorithm results, to group similar and identical strings into **contention sets**
   - **Step 3. Similar strings (applied for)** added to contention sets
   - **Step 4. Similar strings due to variant relations (identified via Hybrid Model)** added to contention sets
7. ICANN communicates the results of the String Similarity Review, including contention sets

IE + EE + Dispute Resolution

8. IE, Extended Evaluation (EE), and Dispute Resolution continue. Some applications may not pass certain elements of the review process, **which may alter the contention sets**
   - **Step 5. Successful String Confusion Objection add strings** to contention sets
   - **Step 6. Strings that fail to pass certain elements of the review are removed** from contention sets

String Contention

9. Is the applied-for gTLD in a contention set?
   - Applicants are encouraged to **self-resolve** string contention anytime prior to the contention resolution process
10. Have one or more community-based applicant(s) elected community priority?
11. Community priority evaluation
12. Auction: one or more parties proceed to next stage
13. Does one clear winner emerge?

Transition to Delegation

14. Applicant enters Transition to Delegation Phase