

# Task Force on Arabic IDNs



**Sarmad Hussain**

Center for Language Engineering  
Al-Khwarizmi Institute of Computer Science  
University of Engineering and Technology  
Lahore, Pakistan

# Recent Community Work on Arabic IDNs

- Very active Arabic script community; time line of recent Arabic script community work
  - 2008: Arabic Script IDN Working Group (ASIWG)
  - 2009: ICANN IDN ccTLDs Fast Track Program initiated
  - 2010: Arabic Language in Internet Domains (RFC 5564)
  - 2011: Arabic Script Case Study for IDN Variants in the Root
  - 2012: User Experience Study on IDN Variants in the Root
  - 2013: Task Force on Arabic Script IDNs by MESWG
- Significant expertise in a variety of relevant domains: Unicode, DNS, IDNA and Registry and Registrar Operations
  - experience from SLD deployments in ccTLDs
  - experience from IDN ccTLD deployments

# Community Driven Way forward: **Task Force on Arabic Script IDNs**

- Creation and Oversight by community based MESWG
- Objectives: a holistic approach
  - Arabic Script Label Generation Ruleset (LGR) for the Root Zone
  - Second level LGRs for the Arabic script
  - Arabic script Internationalized Registration Data
  - Universal acceptability of Arabic script IDNs
  - Technical challenges around registration of Arabic IDNs
  - Operational software for registry and registrar operations
  - DNS security matters specifically related to Arabic IDNs
  - Technical training material around Arabic script IDNs

# Task Force on Arabic Script IDNs

- Membership open, community based
  - Details and interests of members posted by MESWG
  - Discussions publicly archived
  - Details at <http://lists.meswg.org/mailman/listinfo/tf-aidn>
- 
- Open Call for Participation in August-Sept. 2013
  - Announced at Arab IGF II in Algeria on 2<sup>nd</sup> Oct. 2013
  - 20 members from 15 different countries
  - Covering a diverse set of languages which use Arabic script from East Asia, South Asia, Middle East, North Africa, Sub-Saharan Africa

# Label Generation Ruleset for the Root: **Work from Arabic Variant Issues Project**

- Character set

- Ranges

- Arabic: U+0600 – U+06FF
    - Arabic Supplement: U+0750 – U+075F
    - Arabic Extended A: U+08A0 – U+08FF

- Decisions

- What should be included at this time?
    - What should be excluded for all time?
    - What should be decided later?

# Work from Arabic Variant Issues Project

- Character Level Variant Decisions
  - Guiding Principles for Root LGR
  - Analysis of character set included
    - 120+ cases of visually same or similar Arabic script characters identified by case study team
  - Definition of character level variant rules

# Typology of Arabic IDN Variants

- Same
  - Identical U+06CC (ى)/U+0649 (ى)
  - In Context U+06A9 (كِبى)/U+0643 (كِبك)
  - Normalization U+0632 (ز)/U+0631+U+06EC (ر + ّ = ز)
- Similar
  - Character U+06AA (ك)/U+06A9 (ك)
  - Diacritic U+062A (ت)/U+067A (ث)
- Different
  - Shape U+0629 (بة)/U+06C3 (بّة)
  - Character U+0629 (ة)/U+06C1 (ه)

# Label Level Variants

- Label level variant rules
  - Cannot mix Arabic KAF and Persian/Urdu KAF, etc.
  - Other constraints?



شكر يه

Thank you