

# First Global Workshop on Arabic Script Representation in IDNs

Arabic Script IDN Working Group (ASIWG)

April 2, 2008

Dubai, U.A.E.

Sponsored by PIR, United Nations, Afilias

Presented by Ram Mohan ([rmohan@afilias.info](mailto:rmohan@afilias.info))



# Initial Goals

- Establish a framework for implementation of IDNs in Arabic Script
  - **Requires** development of language tables and related Variants (confusingly similar characters)
- Consider, at a minimum, technical, linguistic and policy matters in the development of a Unified Arabic script table.
- Participants attempt to be inclusive with respect to languages that implement the Arabic script throughout the world.
  - Open membership to qualified contributors
- Conform to IETF IDNA Protocol and IDN Guidelines

# Founding Members

- Arabic Language
  - Raed Ibrahim Al-Fayez, Abdulaziz H. Al-Zoman (Saudi Arabia, Arabic language background)
  - Manal Ismail, George Victor (Egypt, Arabic language background)
- Urdu/Sindhi/Pashto
  - Sarmad Hussein (Pakistan, Urdu/Sindhi/Pashto background)
- Farsi
  - Alirezah Saleh, Siavash Shahshahhani, Behdad, Behnam Esfahbod (Iran/Farsi language background)
- Michael Everson (Linguist, Unicode)
- Tina Dam, Baher Esmat (ICANN)

## Organizers:

- Ayman El-Sherbiniy (UN-ESCWA)
- David Maher, Alexa Raad (PIR - .ORG)
- Ram Mohan (Afilias - .INFO)

# Language Group/Script Tables

**Script Level: One Table (Unified Arabic Script Table)**

**Unified Arabic  
Script Table**

**Language Group Level: A Table for each Language, Plus Variant Tables**

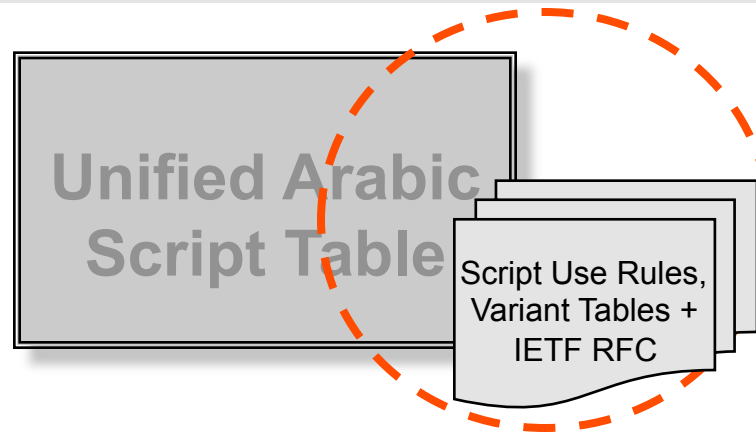
**Arabic Language  
Table**

**Farsi Language Table**

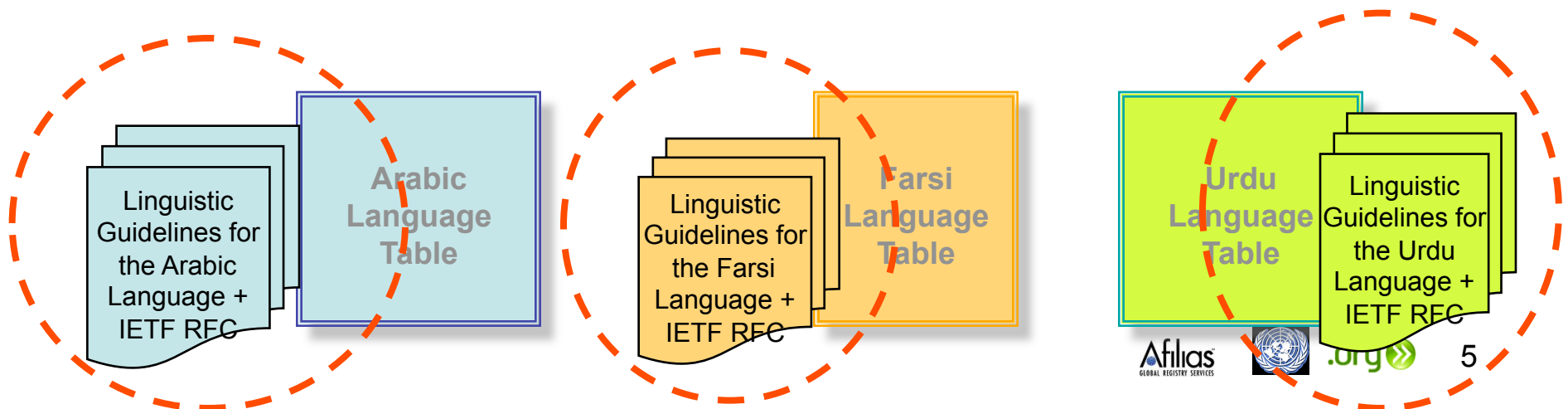
**Urdu Language  
Table**

# Guidelines, rules, RFCs towards IDN TLD Implementations

**Script Level: One Table (Unified Arabic Script Table)**



**Language Group Level: A Table for each Language, Plus Variant Tables**



# Initial Outcomes

Consideration of a *layered* model to address issues with regard to scripts

Level 4	Application/User Level
Level 3	Registry/Language/Policy Level
Level 2	Unified Arabic Script Level
Level 1	IDNA Protocol Level

# Level 1: IDNA Protocol Level

<b>Level 4</b>	<b>Application/User Level</b>
<b>Level 3</b>	<b>Registry/Language/Policy Level</b>
<b>Level 2</b>	<b>Unified Arabic Script Level</b>
<b>Level 1</b>	<b>IDNA Protocol Level</b>

Level 1 deals with for the IDNA protocol level with an impact on Arabic Script.

# Level 2: Unified Arabic Script Level

<b>Level 4</b>	<b>Application/User Level</b>
<b>Level 3</b>	<b>Registry/Language/Policy Level</b>
<b>Level 2</b>	<b>Unified Arabic Script Level</b>
<b>Level 1</b>	<b>IDNA Protocol Level</b>

Level 2 deals with the Unified Arabic Script, with an intent to be inclusive of the languages that use the script

# Level 3: Registry/Language/Policy Level

Level 4	Application/User Level
Level 3	Registry/Language/Policy Level
Level 2	Unified Arabic Script Level
Level 1	IDNA Protocol Level

Level 3 deals with language implementation issues, usually at the level of a Registry operator, and includes policy and/or business rule decisions that affect how IDNs are registered and accessed. Further restrictions on the character set to be used may occur at this layer, and this layer allows some differentiation

# Level 4: Application/User Level

<b>Level 4</b>	<b>Application/User Level</b>
<b>Level 3</b>	<b>Registry/Language/Policy Level</b>
<b>Level 2</b>	<b>Unified Arabic Script Level</b>
<b>Level 1</b>	<b>IDNA Protocol Level</b>

Level 4 deals the software application or user level, and includes display, rendering and user interface decisions that affect how IDNs are accessed by Internet users (both registrants and users of the IDN domain).

# Outcome 1

- Level 1: IDNA Protocol
  - **Outcome:** The group reviewed the current set of DISALLOWED and controversial code points for the Arabic Code Block and have arrived at a set of recommendations that will be sent to the IDNA Protocol list at the IETF for further work.

# Outcome 2

- Identification of Visually Confusable characters
  - Language based approach
    - Arabic, Farsi, Urdu, Sindhi, Pashto represented at the meeting
  - Begun identifying a set of variants that help differentiate between visually confusable characters

# Outcome 3

- Creation of variant tables and guidelines for the use of the Arabic script in specific languages
- Plan is to submit these as either Informational RFCs or Best Current Practice (BCP) RFCs to the IETF

# Structure

- Open Working Group
  - Decisions based on rough consensus
  - Members expected to provide substantive input into the WG, and contribute experts to the group
    - Members fund own participation, to the extent possible
  - Desire participation from:
    - Registries implementing Arabic script
    - Linguistic community experts
    - DNS/Standards experts

# Meeting Plans

- Meet in person *and* in region approximately 2 to 3 times in 2008
  - June/July 2008
  - Oct/Nov/Dec 2008
- Attempt to coincide with other regional/global Internet meetings, to the extent feasible

Mailing List: [ldna-arabicscript@lists.irnic.ir](mailto:ldna-arabicscript@lists.irnic.ir)

Archives: <http://lists.irnic.ir/mailman/listinfo/idna-arabicscript>