IDNA Review and Status

John C Klensin
APNIC
Beijing, 25 August 2009

IDNA History

- Initial version of Standard

  - Approved and published 2003

  - Widely deployed and used in some parts of the world

  - Decision to use in a zone rests with the zone administration

- Issues

  - Exclusion model

  - Some language difficulties

  - Unicode 2.3

Problem Analysis

- Confusion – some deliberate

  – Registries and ad hoc browser decisions the only defense

- Exclusion versus inclusion models

- New Unicode versions

- Mapping confusion

  – Visually confusable characters

  – Other types of confusion

  – Information-losing mapping confused many

- Characters other than letters and digits

IDNA2008

- Work started in design team, mid-2007
  - Unicode independence
  - Inclusion
  - Rule-based, not algorithms
  - Some fixes
- IETF WG established early 2008
  - Assumed about six months
  - Now about 18 and continuing

Why the Delay?

- Inclusion requires making more decisions than "permit anything"
  - Desire to tune to perceived needs of particular languages
  - Unicode not obviously optimal for IDN use
- Backward compatibility issues, particularly
  - Preservation of IDNA2003 Mapping
  - Long debates about a few different character interpretations
- Other agendas

Good News for a Change

- Manila APNIC: No prediction about completion date

- Beijing APNIC: Expect IETF Last Call in September

- Things could still go wrong

  - If all goes smoothly, possible approval in October
  - Or....

Transition 1

- Some labels are valid (but ill-advised) in IDNA2003 but invalid in IDNA2008
  - Registrants will need to migrate legitimate uses (if any)
  - Conforming applications will enforce
- Two characters (0.002%) change interpretation
- Mapping limited and advisory
  - U-labels and A-labels are symmetric
  - Only U-labels and A-labels in IRIs/URIs etc
- New rules and tables

Transition 2

- Registries need policies
  - Most already have them
- Work needed to support new characters
  - Particularly those explicitly banned or ignored in IDNA2003
  - Maybe some need for variants and sunrise, depending on script and usage.

IDNA2008 Will Still Not Solve

- Connectivity problems

- Keyboard problems

- Complex policy problems

  - Internationalization tends to increase such problems

- Language-specific issues within a script

  - DNS doesn't know about languages

  - Label strings are generally too short to establish language context (mnemonics are impossible)

But...

- IDNA2008 will provide a better foundation for working on those problems than IDNA2003

- We may be closer to being ready to address them