I18N for RDAP

Andy Newton
(with guidance from Peter Saint-Andre)

WEIRDS WG, IETF 86
Lookups vs Searches

• Lookups
  – draft-ietf-weirds-rdap-query
  – draft-ietf-weirds-json-response-02
  – User expects zero or one results, typically has a semi-unique key to use with the query

• Searches
  – draft-hollenbeck-weirds-rdap-search
  – draft-zhou-weirds-rdap-restful-search
  – User expects zero or many results, typically has only partial information for which to start the query
I18N in Lookups

• From Atlanta
  – Added JSON 'lang' attribute for queries
  – No ability to express desired lang in query
  – Registries either have the data or they don't

• U-Labels in query?
  – Not a good idea
  – Domain lookups are about what winds up in the DNS zone files
  – draft-ietf-weirds-json-response should be more explicit about domains being A-label
Searches

• Semantics of partial matching with respect to Unicode
  – How does that work?
  – Canonicalization of inputs?
  – Partial matches against Unicode combining characters, etc…
Does a search for "exae" yield "exæ"?

Does a search for "exa*" yield?
exä
exå
exà
exæ
exå
exâ
Search is Hard!

• Or is it...

• Canonicalization
  – Solved for us with domain names by IDNA
  – Solved for us with proper names by PRECIS and (something similar to ) draft-ietf-precis-nickname

• Partial Match
  – If the user wants exä exå exæ exå exâ in results they should search on "ex*"
  – User agent could send multiple queries and concatenate and filter appropriately
Throw fruit now!