Activity Since IETF 83

• Added OtherLetterDigits category \{Lt, Nl, No, Me\}

• Completed first draft of codepoint table (note: manually generated!)

• Specified two more IANA registries, for base classes and subclasses

• Recommended “SubclassBaseClass” format for subclass names
Processing Order

• IDNA2008: validation, then normalization
• PRECIS (now): normalization, then check string validity
• Some inconsistent results (e.g., in Hangul)
• Change order in PRECIS to be consistent with IDNA2008?
Space: Problem

- All whitespace characters (even U+0020) are prohibited in NameClass
- NameClass is used for usernames
- But in SASL (and LDAP?), usernames might be full names with spaces (e.g., “Peter Saint-André”)
- (Note: we don’t know how common this is in deployments of SASL)
Space: Concerns

• In some locales, input methods don’t generate U+0020 (e.g., space bar => Zero Width Joiner)

• Spaces are not always perceptible in user interfaces

• Many characters are confusable with U+0020 (e.g., U+180E, U+2028, U+2029, U+2000..U+200A)

• Basic concern seems to be violating the principle of least user surprise
Space: Solutions?

- PRECIS base class “CompoundNameClass” that allows U+0020 (but not any other Zs codepoint)
- Application-layer re-use of NameClass, e.g., compound = name [I*(SP name)]
- Escaping, e.g., “Peter%20Saint-Andre”
- Tell SASL and LDAP to stop supporting spaces
- Other?
Remaining Tasks

• Close the processing order issue
• Find a solution to the space problem
•Finish code for automated generation of the codepoint table
• Check usage in XMPP, SASL, etc.
• Gather more implementation experience