

Internationalization of Resource Registry

Piotr Strzyżewski

Silesian University of Technology, Computer Centre



Reason

- ❑ Various languages and alphabets used natively in RIPE Service Region
- ❑ Even more languages and alphabets used by number of inhabitants in RIPE Service Region
- ❑ Some alphabets use latin characters and some not
- ❑ Using names, addresses, words written using US-ASCII generates problems

Example

- Consider two polish words (and three 😊 meanings):
 - Polka – Pole, Polish woman
 - polka – polka – Czech national dance
 - półka – shelf, rack

Example

- Now, consider two names put into database:
 - Maria (Mary) Polka
 - Maria Półka
- For both of them we have to put „Maria Polka” into database
- People don't like mistakes in their names! ☹

Example

- Now, consider two addresses put into database:
 - ul. (Marii) Polki 1/1 (genitive)
 - ul. (Marii) Półki 1/1
- For both of them we have to put „ul. (Marii) Polki 1/1” into database
- People, post and LEAs do not like such things! ☹



Idea

- Modify the DB software and introduce internationalization mechanisms
- Initial idea is to introduce some idn fields in form of „name-idn” as an extension to currently used:
 - person-idn
 - address-idn
 - remarks-idn
 - etc



Problems

- Which code page should be used? UTF-8?
- Does automatic code page translation should be considered?
 - If yes, which code pages? ISO 8859 family? Windows code pages? IBM code pages?
- Does automatic US-ASCII conversion should be considered (when possible)?
- Whois client should be able to ask for limited or full data.
- Interoperability problems with other databases



Pros and cons

- (+) Data will be more accurate

- (-) Some problems have to be issued and solved



Questions?