

Multilingual and Other Localization Considerations for the OOR

Ken Baclawski
Northeastern University
Jan-Eric Litton
Karolinska Institutet
Neil Sarkar
University of Vermont

Background

- Data sources span regional and international boundaries
 - Different natural languages
 - Regional differences within the same natural language
 - Exact equivalences are not always available
- Terminology evolves over time
- Terminology is closely tied with legal requirements and policies

Biobanks

- Repositories of biological materials
- Associated clinical information include
 - Source (donor)
 - Type of sample
 - Genotypes
 - Medical records
 - Process used to obtain the sample
 - Policies for use of the sample
 - Donor specified restrictions
- Joint work with Jan-Eric Litton, Karolinska Institutet, Stockholm

Disease Reports

- Diseases do not respect national boundaries.
- Reports relevant to human and animal diseases are in many languages.
- Terminology variation by region and time
 - Diseases
 - Species
 - Habitats
- Joint work with Neil Sarkar, University of Vermont

Repository Requirements

- Multilingual browse and query capability
 - Integration of multiple languages and regional variations
 - Integration of terminology that changes over time
- User-friendly interface for suggestions and feedback from the community
 - Terminology
 - Processes
- Requirement and policy checking support