Subject authority control supported by classification: the case of National Library of the Czech Republic

Marie.Balikova@nkp.cz
Outline

• Brief characteristics of subject access and authority files
• Where subject authorities are applied
• Strengths and weaknesses of traditional authority files in current online environment
• Application of subject authorities in mapping/harmonization process of terminological resources in INTERMI project
• Conclusion
Subject access in Czechia in 1990's

Two important questions:

• Is it possible to create a universal subject access tool?
• How might subject authority control and classification schemes improve subject access of information resources?

We based on several assumptions

• As a library with universal collections we needed to develop a universal subject access tool because it was not possible to create a universal system by merging domain specific control vocabularies or thesauri.
• Subject authority control is necessary because it represents a fundamental prerequisite for the creation of controlled access points.
• Classification schemes as hierarchical and universal knowledge organization systems can support formation of hierarchical structures and adapt the subject access to specific groups of users.
Subject authority control in Czechia

The most important achievement has been the launch of the subject authority file as a controlled vocabulary (2000). It serves as a base platform for creating other user-friendly organising tools such as

– Conspectus categorisation scheme (CCS)
– scheme for Uniform Information Gateway (UIG)
– scheme for subject information gateways (SIG)
– Topical Map of National Library Collections
– Topical authorities online

• It serves as a basis for mapping/harmonization of terminological resources used in memory institutions as well
Czech National Authorities

- Name authorities were originally created according to AACR2 rules, since May 2015 RDA rules
- Subject authorities have been created according to NLCR rules (based on LCSH rules)
Czech Subject Authority files - CZENAS

represent a structured controlled vocabulary systematically identifying the basic semantic relationships (equivalence, hierarchical and associative) between terms in natural language that is designed for both, post-coordinated and pre-coordinated approach

Topical file is a universal/general controlled system which includes terms not only for general concepts, and topics but for events (historical included) as well

Marie Balíková, NL CR
Example of topical authority record

The subject authority record is a set of information about indexing term.

Each authority record contains the established form of the indexing term, *see from* references, *see also* references, equivalent UDC notation, equivalent Conspectus category, English equivalent and notes.

The record also contains coded data that describes characteristics of the record.
Czech Subject Authorities in Conspectus Categorization scheme

- **Conspectus method** is an international standard whose primary aim is the coordinated building of library holdings and access to them through content characteristics.

- The Conspectus method was developed for use in libraries organized either using the Library of Congress (LCC) classification or the Dewey Decimal Classification (DDC) structures.

- It was necessary to accommodate the Conspectus scheme to the local practice defining the concordances between DDC and UDC numbers and the corresponding subject terminology in English and Czech.
Where is the Conspectus categorization scheme applied

The categorization scheme for the needs of Conspectus comprises three hierarchical levels, namely:

- 24 basic groups, the so-called Conspectus Divisions
- 584 subordinated categories, the so-called Conspectus groups/categories
- topical terms of the subject authority file (CZENAS)

- the most important characteristic of the Conspectus scheme - access to information resources through content characteristics expressed by combinations of UDC notations and verbal expressions - was used when creating the
- Topical Map of the Czech National Library Collections
- Topical Authorities Online
- subject-oriented universal gateway (UIG – Uniform Information Gateway), subject information gateways (KIV /Library and Information Science/, MUS, TECH, ART)
Topical Map of National Library Collections: user-friendly tool for publishers ...

National Library of the Czech Republic - Topic Map of Library Collections

Select all categories

Conspectus subcategories within selected categories display:

- in alphabetical order (without DDC notations)
- according to DDC notations

Marie Balíková, NL CR
Topical Authorities Online: access to bibliographic records using topical authorities

Marie Balíková, NL CR
The purpose of UIG and subject gateways created within the UIG portal is to allow users to browse Czech and foreign resources (library and union catalogues, full-text databases etc.) in parallel, from the respective semantic domains by means of a single search interface and from a single entry point.
Application of subject authorities in mapping/harmonization process of terminological resources in INTERMI project
Interoperability in Memory Institutions project

Goals of INTERMI project

• to develop an infrastructure for creation a knowledge-based model of the cultural heritage content
• to create conceptual model and ontology for making Czech national cultural heritage content available on the web for professionals and general public as well
• to assess the applicability of traditional library standards and systems for access to the national cultural heritage in Czechia
• to support collaboration among memory institutions communities (GLAM communities)
The identification of information objects – entities - on the metadata level is based on both name and subject authorities.
Strengths of traditional authority files

• **standardized access points** (preferred and variants forms) are established and shared
• **best practices** are well documented and shared
• **cooperative projects** among libraries and between National Library and some memory institutions, as
  – *Cooperative creation and use of National Authority files*
  – *Museum Authorities*
  – *Register of Fine Art collections*
Weaknesses of traditional authority files

• encoded in MARC format
• record/document based system, not data based system
• lack of data granularity
• entities of the same type in different authority files
• little information in authority records
• lack of term describing type of corporate body, e.g.
  „s.r.o. společnost s ručením omezeným“ – „ltd/llc limited (liability) company“
Role of National Authority files in INTERMI project

• preferred and variant forms of entities appellations shall be used
• information that provides context to entities in authority records shall be accepted and entered in structured form
• the tendency to make decisions based purely on the format (MARC format) should be avoided whenever possible
• the INTERMI conceptual model shall be based on entities (not on authorities)
• proprietary conceptual model and data structure format have to be created
Technologies of Semantic Web applied in INTERMI project

- **URI** (uniform resource identifier) used for identification of objects and abstract concepts;
- **RDF and RDF Schema** for representation of information using abstract metadata model, RDF allows to registration of simple statements with triplets (subject – predicate – object);
- **OWL** – language for ontology creation;
- **SPARQL** – query language.

Various standards related to Linked Data concept and Semantic Web were considered to find one or more suitable standards for data presentation.
- **SKOS** – for representation of thesauri – can be applied on entities from class “General concept”,
- **FOAF** – ontology for describing persons, activities and relations – can be applied on entities from class “Person/creature”.

Marie Balíková, NL CR
Mapping/harmonization process of terminological resources used in memory institutions depends on

• the choice of a **controlled system** of terms that may serve as a **basis** for evaluating terms from other terminological resources

• the choice of a **tool** for formal representation of terminological resources (**SKOS**)  

• the choice of **methods** applied in process of mapping/harmonization, such as lexical-based, concept-based and instance based-mapping
Topical authorities – basis for mapping/harmonization of terminological resources used in memory institutions

CZENAS represents a general set of terms with content covering all fields at general and specific level as well

Transformation needed in order to

• define the meaning of topical terms exactly and systematically by adding qualifiers, UDC notations and scope notes, so that the topical terms are able to represent isolated well defined concepts,

• to express hierarchical structure exactly to be complete and consistent,

• to use permanent unique identifier which shall be language independent to avoid the necessity to change ID in conjunction with the change of preferred form applied in description of the concept,

• to express topical terms file in machine-understandable way within the framework of the Semantic Web, when using RDF application SKOS; it enables data to be linked and merged with other RDF data by Semantic Web applications.
Example of common graphs method used in concept-based mapping of terminological resources

Concept-based mapping is applied when terminology of specialized thesauri and controlled vocabularies is compared. The role of UDC notations as language independent codes linked to the verbal expressions is evident, more research is required.
Conclusion – what we have learned

effective subject access cannot exist without standardized access points, without subject access control (Tillet, Gorman...) in addition to subject domain aimed thesauri, there is a need to have available a universal controlled access tool e.g. national subject authority files as well, such a universal controlled access tool can be created with the support of universal classification scheme only, the application of UDC in current environment depends on the availability of the system online, therefore we appreciated the building of web service UDC online which is available on the address http://www.udc-hub.com/
Universal Classification Scheme has been chosen because it
covers all subjects
supports identification of terms
identifies the position of the term in the terminology systems of individual disciplines
provides context to the search terms
supports the formation of hierarchical structures
enables language independent coding
supports browsing and navigation in large collections of information resources
supports mapping/harmonization of terminological resources used in memory institutions
Thank you!

Czech National Library Subject Authorities Team