# LINKING THE WEB OF KNOWLEDGE:
## UDC's role in bridging information services and resources

Aida Slavic
UDC Consortium, The Hague
aida.slavic@udcc.org

Marie Balikova
Czech National Library
Marie.Balikova@nkp.cz

UDC CONSORTIUM

# OUTLINE

- Facts and figures (UDC 1992-2011)
- UDC applications today: scope and potentials
- How to make classification work in knowledge discovery
- UDC in linking legacy collections and web resources
- UDC as linked data (Simple Knowledge Organization Systems - SKOS) - development and issues

# FACTS AND FIGURES (UDC 1992-2011)

- Translated in over 40 languages, used in over 130 countries:
  - bibliographies and bibliographic databases
  - libraries (also some museums, archives)
  - digital collections, web portals, alerting services

- Annually updated and distributed as a file:
  18 versions/'editions' since 1992
  - 1992: 60,000 classes - 2011: 69,000 classes
  - 10,000 classes cancelled
  - 19,000 new classes added

# LATEST DEVELOPMENTS

## Supporting sharing and use of UDC world-wide

- 2009 UDC Summary – 2,400 classes released under Creative Commons Licence (45 languages)
- 2010 data services helping publishers to align their old editions with the latest version of the UDC and hosting of translation database online
- 2010-2011 Czech, English, Spanish, Dutch, German, Portuguese, Croatian, Bulgarian, Swedish translations imported in the UDC MRF database
- 2011 (November) UDC Summary 26 languages published as SKOS (XML/RDF) http://udcdata.info/udcsummary-skos.rdf to support m2m use of UDC

# UDC APPLICATIONS: SCOPE AND POTENTIALS

- Cross-language linking of concepts, i.e. managing link between concepts and language

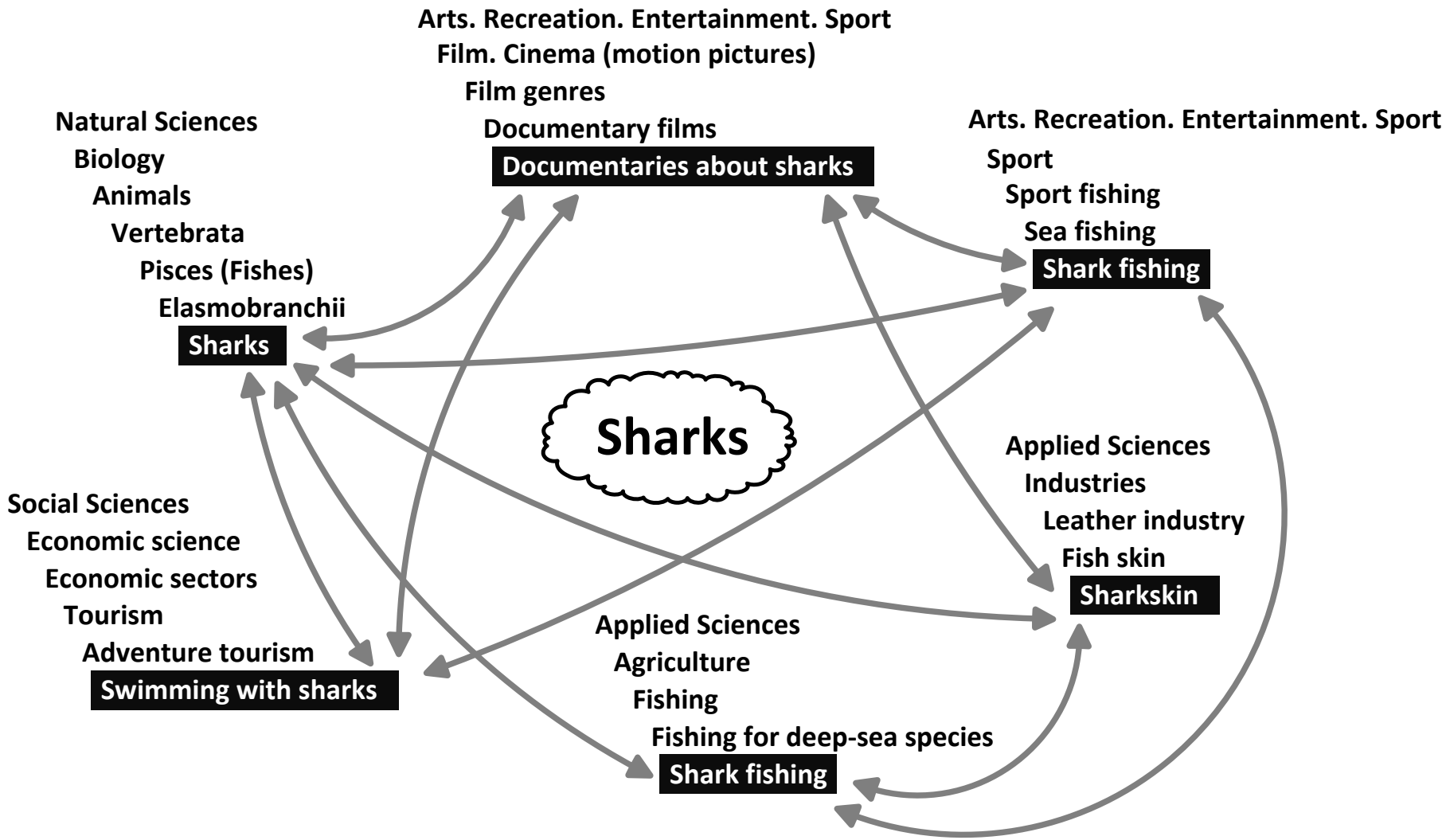
  =512.16 Jižní skupina turkických jazyků
  Южная группа тюркских языков [Russian]
  तुर्की भाषाओं का दक्षिणी समूह [Hindi]
  Թուրքական լեզուների հարավային խումբ [Armenian]
  Νότια ομάδα των Τουρκικών γλωσσών [Greek]
  突厥南部语 [Chinese]
  দক্ষিণস্থ শ্রেণির তুর্কি ভাষাসমূহ [Bengali]
  チュルク語南部群 [Japanese]
  ಟರ್ಕ್ ಭಾಷೆಗಳ ದಕ್ಷಿಣ ಭಾಗದ ಸಮೂಹ [Kannada]

- Hierarchies: graphic knowledge presentation, browsing knowledge space (supporting interactive user behaviour)

- Linking concepts

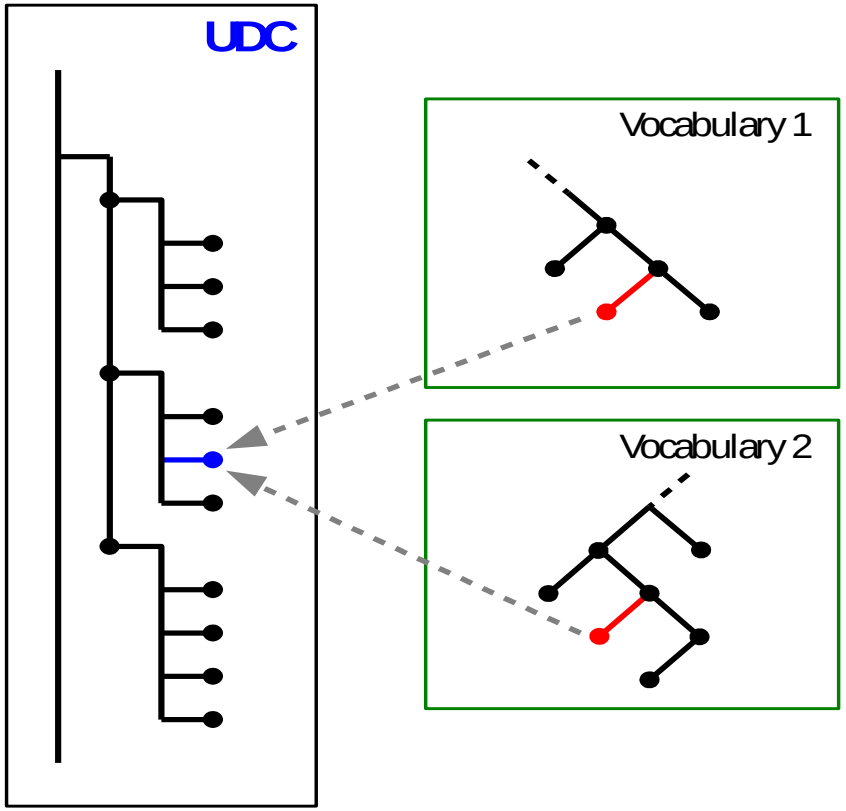
  'fish' in zoology, in sport, in cooking, in food industry, in animal husbandry

# LINKING CONCEPTS ACROSS KNOWLEDGE



# UDC AS A SWITCHING LANGUAGE

| DDC | UDC |
| --- | --- |
| 536 Heat | 536 Heat. Thermodynamics |
| Chaleur | Chaleur. Thermodynamique |
| Теплота | Тепло. Термодинамика |
| 热、热力学、 | 热 |



# CREATING THE WEB OF KNOWLEDGE

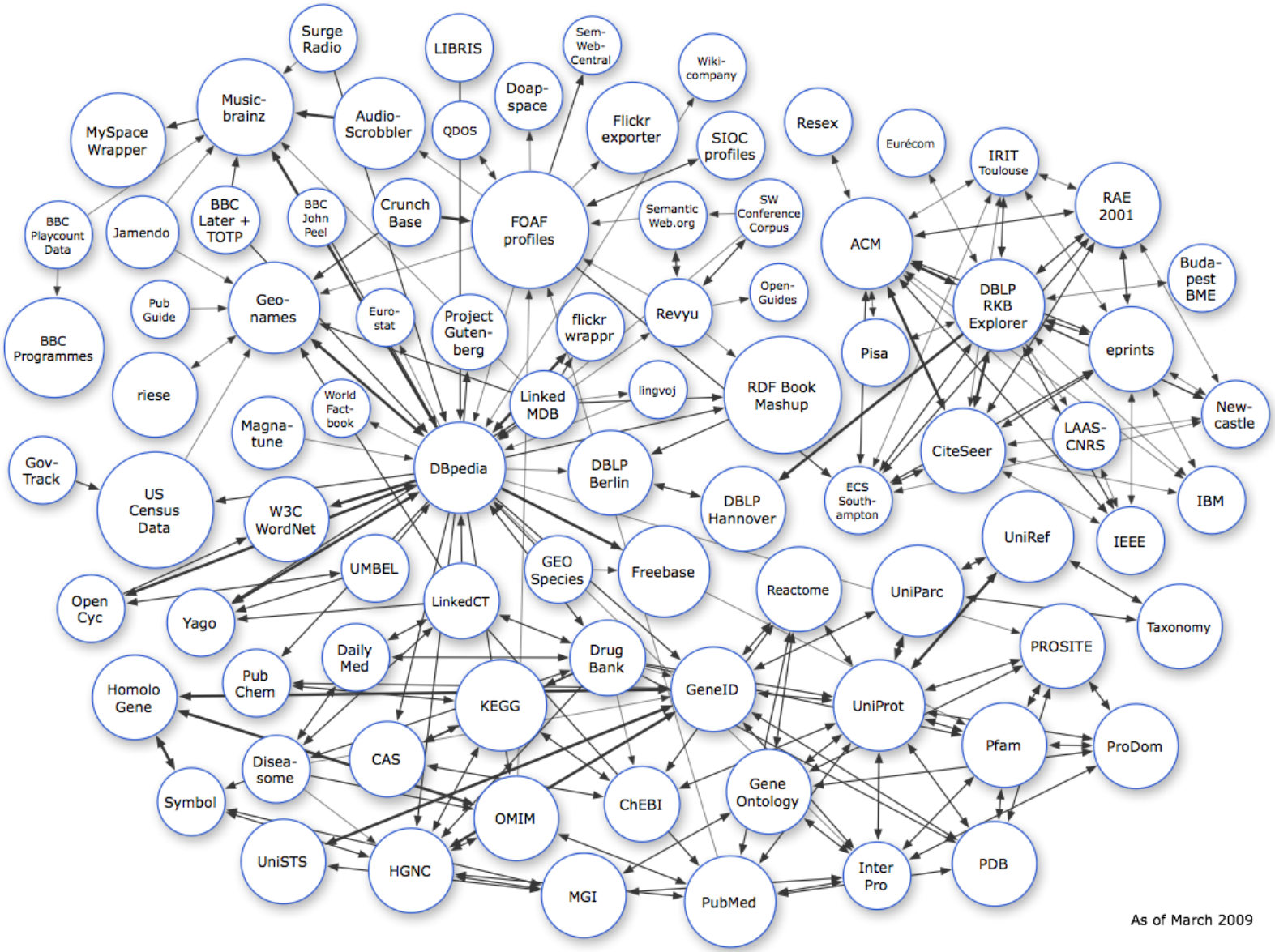
Classifications can be used on the Web to:

- Improve and enrich semantics and access points in the retrieval of information
- Enable information discovery across collections and languages

Two requirements:

- publishing classification: open access to classification vocabulary for m2m processing
- publishing library catalogues: open access to collections and collections' metadata for m2m processing

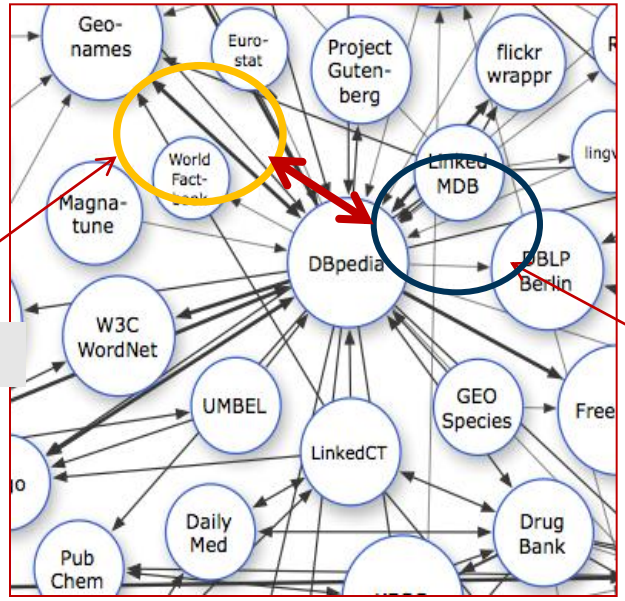
# LINKED DATA


Surge Radio
LIBRIS
Sem-Web-Central
Wiki-company
Doap-space
Music-brainz
Audio-Scrobbler
QDOS
Flickr exporter
SIOC profiles
Resex
Eurécom
IRIT Toulouse
MySpace Wrapper
BBC Playcount Data
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
Semantic Web.org
SW Conference Corpus
ACM
RAE 2001
Buda-pest BME
BBC Programmes
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Revyu
Open-Guides
DBLP RKB Explorer
Pisa
eprints
riese
World Fact-book
Linked MDB
lingvoj
RDF Book Mashup
New-castle
Magna-tune
Gov-Track
US Census Data
DBpedia
DBLP Berlin
DBLP Hannover
ECS South-ampton
CiteSeer
LAAS-CNRS
IBM
W3C WordNet
UMBEL
GEO Species
Freebase
UniRef
IEEE
Open Cyc
Yago
LinkedCT
Reactome
UniParc
Taxonomy
Daily Med
Drug Bank
PROSITE
Homolo Gene
Pub Chem
KEGG
GeneID
UniProt
Pfam
ProDom
Disea-some
CAS
Gene Ontology
Symbol
OMIM
ChEBI
UniSTS
HGNC
MGI
PubMed
Inter Pro
PDB
As of March 2009

# M2M ACCESS - LINKED DATA - SEMANTIC WEB

connecting collections of data by programmes (machine-to-machine)

XML/RDF presentation relies on unique identification of resources (URI) pointing to one another




XML/RDF export

XML/RDF export

COLLECTION CATALOGUES

UDC

# UDC LINKED DATA - DEVELOPMENT TO DATE

- UDC Summary, 2400 skos:concepts published as a Simple Knowledge Organization System (SKOS ) in November 2011
  - independent URI: http://udcdata.info/
  - 26 languages
  - static XML/RDF export
  - official release & dynamic export 12 December 2011

- Relevance of UDC Summary
  - good subject coverage, carefully selected in order to support training, testing, and the majority of information exchange purposes
  - notes, examples, references, alphabetical index, mappings to other systems
  - administrative data: introduction date, revision date, notation and concept history (links to cancelled notations)

# Universal Decimal Classification

summary

45 languag

Czech (Čeština) [cs]

Czech (Čeština) [cs]
Danish (Dansk) [da]
Dutch (Nederlands) [nl]
English (English) [en]
Esperanto (Esperanto) [eo]
Estonian (Eesti) [et]
Finnish (Suomi) [fi]
French (Français) [fr]
Galician (Galego) [gl]
Georgian (ქართული) [ka]
German (Deutsch) [de]
Greek (Ελληνικά) [el]
Hindi (हिन्दी) [hi]
Hungarian (Magyar) [hu]
Indonesian (Bahasa Ind.) [id]
Irish (Gaeilge) [ga]
Italian (Italiano) [it]
Japanese (日本語) [ja]
Kannada (ಕನ್ನಡ) [kn]
Latvian (Latviešu) [lv]
Lithuanian (lietuvių) [lt]
Malayalam (മലയാളം) [ml]
Marathi (मराठी) [mr]
Norwegian (Norsk) [no]
Polish (Polski) [pl]
Portuguese (Português) [pt]
Romanian (Româna) [ro]
Russian (Русский) [ru]



zobrazit vše | skrýt vše

ZÁKLADNÍ SCHÉMA

- VŠEOBECNÉ POMOCNÉ ZNAKY A JEJICH ČLENĚNÍ
  - + Přiřazení (Tabulka 1a)
  - / Rozšíření (Tabulka 1a)
  - : Vztah (Tabulka 1b)
  - :: Pevné (nezaměnitelné) pořadí (Tabulka 1b)
  - [] Podřazení (Tabulka 1b)
  - * Pomocné členění s použitím notace ze zdrojů mimo MDT (Tabulka 1h)
  - A/Z Přímé abecední členění (specifikace) (Tabulka 1h)
- POMOCNÉ TABULKY
  - =... Všeobecné pomocné znaky jazyka. Tabulka 1c.
  - (0...) Všeobecné pomocné znaky formy. Tabulka 1d.
  - (1/9) Všeobecné pomocné znaky místa. Tabulka 1e
  - (=...) Všeobecné pomocné znaky plemen, etnických skupin a národností. Tabull
  - "..." Všeobecné pomocné znaky času. Tabulka 1g
  - -0... Všeobecné pomocné znaky obecných charakteristik: vlastností, materiálů,
- Hlavní tabulky
  - 0 VĚDA A VĚDĚNÍ. ORGANIZACE. VĚDA O POČÍTAČÍCH. VÝPOČETNÍ TECHNIKA. I

záznam získ
pokliknutím
znak/symbo
vlevo

# THE FIRST RELEASE AS LINKED DATA

The first stage contains the following UDC data:

- UDC number (notation) ➔ skos:notation
- class identifier (URI) ➔ skos:Concept
- broader class (URI) ➔ skos:broader
- caption ➔ skos:prefLabel
- including note ➔ skos:note
- application note ➔ skos:note
- scope note ➔ skos:scopeNote
- examples ➔ skos:example
- see also reference ➔ skos:related

# UDC SKOS EXPORT (XML/RDF)

```
−<skos:Concept rdf:about="http://udcdata.info/068288">
    <skos:inScheme rdf:resource="http://udcdata.info/udcSummary"/>
    <skos:broaderTransitive rdf:resource="http://udcdata.info/000109"/>
    <skos:notation rdf:datatype="http://udcdata.info/UDCnotation">=162.3</skos:notation>
    <skos:prefLabel xml:lang="en">Czech</skos:prefLabel>
    <skos:prefLabel xml:lang="nl">Tsjechisch</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Checo</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Tchèque</skos:prefLabel>
    <skos:prefLabel xml:lang="sv">Tjeckiska</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Tschechisch</skos:prefLabel>
    <skos:prefLabel xml:lang="hr">Češki</skos:prefLabel>
    <skos:prefLabel xml:lang="ru">Чешский язык</skos:prefLabel>
    <skos:prefLabel xml:lang="pl">Język czeski</skos:prefLabel>
    <skos:prefLabel xml:lang="ro">Ceha</skos:prefLabel>
    <skos:prefLabel xml:lang="cs">Čeština</skos:prefLabel>
    <skos:prefLabel xml:lang="hu">Cseh</skos:prefLabel>
    <skos:prefLabel xml:lang="uk">Чеська</skos:prefLabel>
    <skos:prefLabel xml:lang="hi">चेक</skos:prefLabel>
    <skos:prefLabel xml:lang="et">Tšehhi</skos:prefLabel>
    <skos:prefLabel xml:lang="hy">Չեխերեն</skos:prefLabel>
    <skos:prefLabel xml:lang="pt">Checo</skos:prefLabel>
    <skos:prefLabel xml:lang="ca">Txec</skos:prefLabel>
    <skos:prefLabel xml:lang="el">Τσεχική</skos:prefLabel>
    <skos:prefLabel xml:lang="zh">捷克语</skos:prefLabel>
    <skos:prefLabel xml:lang="bn">চেক</skos:prefLabel>
    <skos:prefLabel xml:lang="sk">Čeština</skos:prefLabel>
    <skos:prefLabel xml:lang="ja">チェコ語</skos:prefLabel>
    <skos:prefLabel xml:lang="bg">Чешки език</skos:prefLabel>
    <skos:prefLabel xml:lang="kn">ಜೆಕ್</skos:prefLabel>
    <skos:prefLabel xml:lang="be">Чэшская мова</skos:prefLabel>
  </skos:Concept>
```

example of the UDC class

=162.3 Czech

[Common auxiliary of language]

# NEXT STAGES

## adding new data

- mappings to classifications and subject heading systems
- development of alphabetical index
- historical and administrative data (to include redirections, cancelled classes)
- documentation, versioning control

## extension of the existing data

- more UDC classes necessary to support information exchange on the Web
- more exports and documentation data for free download
- more languages/translations

# ISSUE 1: OBSOLETE AND NEW UDC NUMBERS

- Enable automatic redirection on the Web from cancelled UDC numbers. UDC MRF database holds data as follows:

| | | |
|---|---|---|
| UDC CLASS NUMBER: | 22 | |
| DESCRIPTION: | The Bible. Holy scripture | |
| REPLACED BY: | 26-23 | Judaism – Scriptures |
| | 27-23 | Christianity - Scriptures |

- SKOS does not offer solution for presenting this kind of data at the moment

- But in RDF, it is possible to use other models in combination with SKOS…
  - Dublin Core for versioning links the two classes using properties in the term namespace : *isReplacedby*; *replaces*

# ISSUE 2: RESOLVE COMPLEX UDC NUMBERS

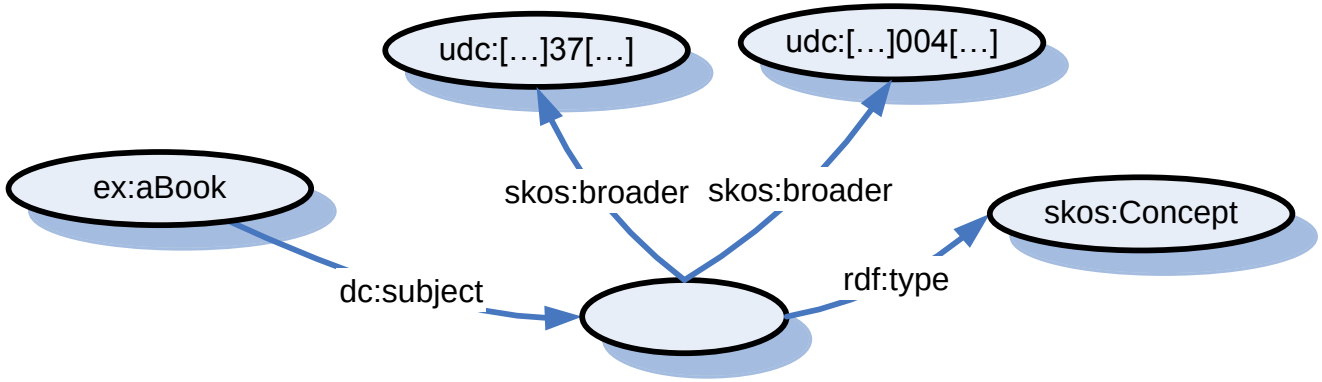
- Complex UDC expressions appear in the process of use and may not appear in the original scheme

| | |
|---|---|
| 37:004 | Application of computers in education |
| 32:37 | Relationships between politics and education |

- Library catalogues or authority data shared on the web contain many pre-combined number and when published as linked data these numbers may have their own URI

- How to link representations between notations from the original schemes and complex subject expressions developed at the point of indexing?

# POSSIBLE SOLUTION: A SUBJECT NODE




- Create a new empty subject node that is connected to the combined concepts
- The node can be represented as anonymous: if library systems do not provide IDs for these collection-specific creations
- This node can be used to relate 37 and 004 classes, expressing that the new anonymous class is more specialized
  - this would allow for finding of the book starting from a query combining the skos:Concepts standing for 37 and 004

# CONCLUSION

## supporting the use of UDC in collections

- regular updates, expansions of vocabularies
- multilingual access
- mappings to other knowledge organization systems
- supporting automation and automated management of subject access and automatic updates of authority files

## for use on the Web (Semantic Web)

- more and better open access to UDC data on the Web
- machine-to-machine access (ontology-like representation) e.g. XML/RDF

# THANK YOU

Useful links:

- UDC Summary
  http://www.udcc.org/udcsummary/php/index.php?lang=cs
- UDC as linked data
  http://udcdata.info/udcsummary-skos.rdf (10Mb!)
- SKOS Simple Knowledge Organization System: Reference
  http://www.w3.org/TR/skos-reference/
- DCMI Metadata Terms
  http://dublincore.org/usage/terms/