An in-depth view behind the scenes: grammar, semantics and management of thesauri

Dr Jeannine Beeken

Discovery Officer: Systems, User Interfaces, Schema

jeannine.beeken@essex.ac.uk

Taxonomy Boot Camp London 2017

17 – 18 October

Olympia London

UK Data Service

UKDS HASSET - CESSDA ELSST

- UKDS: UK Data Service – UK Data Archive, University of Essex

- HASSET: Humanities and Social Science Electronic Thesaurus, monolingual English, UK coverage

- CESSDA: Consortium of European Social Science Data Archives

- ELSST: European Language Social Science Thesaurus, multilingual (13 languages), EU coverage

- Usage: indexing and retrieval, fully integrated data archives infrastructure

UK Data Service

EN extension–EN international–EU Lang

- 13 languages: Czech, Danish, English, Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Slovenian, Spanish, Swedish

- Monolingual mapping: HASSET and ELSST

- Multilingual translations: within ELSST

- 1 Thesaurus Management Tool (in-house)

HASSET EN ⟷ ELSST EN ⟷ CS

ELSST EN ⟷ DE

ELSST EN ⟷ GR

UK Data Service
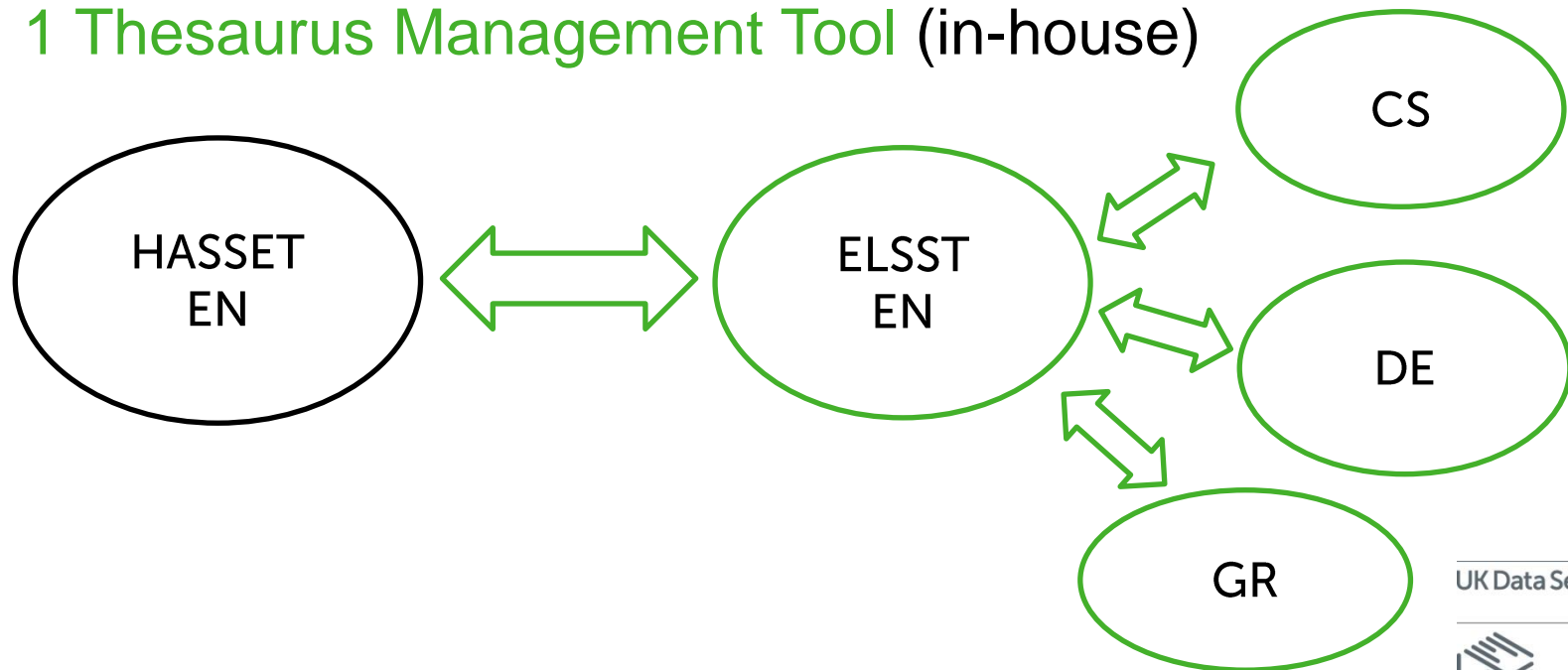
ELSST Participants

FINNISH SOCIAL SCIENCE DATA ARCHIVE

ČSDA Český sociálněvědní datový archiv
Sociologický ústav AV ČR, v.v.i.

Danish Data Archive - DDA
RIGSARKIVET

NSD NORWEGIAN CENTRE FOR RESEARCH DATA

FORS +

SND Swedish National Data Service

EKKE NATIONAL CENTRE FOR SOCIAL RESEARCH

gesis Leibniz Institute for the Social Sciences

AUSSDA AUSTRIAN SOCIAL SCIENCE DATA ARCHIVE

LISER LUXEMBOURG INSTITUTE OF SOCIO-ECONOMIC RESEARCH

ROMANIAN SOCIAL DATA ARCHIVE

VILNIAUS UNIVERSITETAS · 1579 · UNIVERSITAS VILNENSIS
Vilnius University Library

ADP - SOCIAL SCIENCE DATA ARCHIVES
Analyze data! Deposit study! Promote science!
1997 - 2017

UK Data Service

Overview

- Semantic types
- Semantic relationships
- Requirements: behind the scenes/(user interface)
- Rules and constraints: mapping 2 English thesauri
- Editing 2 monolingual thesauri
- Editing multilingual thesaurus (13 languages)

- History/change management
- Suggestions/input management
- Bulk upload

UK Data Service

Semantic types

- Terms = labelled concepts
    - Preferred terms (PT)
    - Use for terms (UF): synonyms of PT
    - Qualifiers in PTs and UFs disambiguate homographs

- Notes
    - Scope note: disambiguate polysemes, sets the semantic scope
    - Scope note source: trusted reference
    - Use note: reference to other terms in thesaurus
    - History note: information about the history of the term
    - Editorial note: Admin/Translators of a specific language

UK Data Service

Example semantic types

Semantic relationships

- Clustering
  PTs and UFs

- Hierarchical: taxonomy, reverse relationship

  - top terms (TT)/broader terms (BT)

  - reverse narrower terms (NT)

  - polyhierarchies: more than 1 BT

- Associative: ontology, symmetrical/reciprocal relationship

  - relational terms (RT)

UK Data Service

Example semantic relations: tree view

Example semantic relations: graph

Example: semantic polyhierarchy

Requirements 1/6

- Compliance standards/integration (behind the scenes) ISO 25964, SKOS/SKOS-XL/XKOS vocabularies, multiple display modes, multilingual, versioning

- User management role-based permission (+slide) (bts)

- KO schemes (bts, UI) concept-based/ terminology-based, create/manage extensible set of semantic relation types/information types/monohierarchies and polyhierarchies (+slide)

- Visual modes (UI) alphabetical, hierarchical, systematic, graphs

UK Data Service

User management: User Roles

Role -> Responsibility ↓	General users VocabularySource: ELSST or HASSET	Viewer/Guest after login VocabularySource: ELSST or HASSET	TI administrator/ Translator with translator rights VocabularySource: ELSST TL	SL administrator/TLs administrator/ content administrator VocabularySource: ELSST SL and TLs/HASSET	Technical Administrator
Concept view: view and search all externally published vocabularies/Concepts and their items present in ELSST Current version SL and TLs/HASSET	√	√	√	√	√
Concept view: view and search all vocabularies/Concepts within the VocabularySource and their items whether published or not. i.e. ELSST Current and Next version SL and TLs/HASSET	x	√	√	√	√
Suggestions view: view, search within the VocabularySource, i.e. ELSST Current and Next/HASSET and edit: add to discussion within the VocabularySource, i.e. ELSST Next SL and TLs/HASSET	x	√	√	√	x
History view: view and search within the VocabularySource, i.e. ELSST Current and Next SL and TLs/HASSET	x	√	√	√	x
Suggest a change: view, search, edit: suggest add/create, suggest update/change, suggest deletion within the VocabularySource, i.e. ELSST Current and Next SL and TLs/HASSET	x	√	√	√	x
TL Translation language dependent: view, edit: add/create, change/update, delete within the VocabularySource, i.e. ELSST Next for TL	x	x	√	√	x
SL and TLs edit: add/create, change/update, delete within the VocabularySource, i.s. ELSST Next SL and all TL/HASSET	x	x	x	√	x
Publish internally ELSST Next SL and TLs/HASSET	x	x	x	√	x
Publish externally ELSST Next SL and TLs/HASSET	x	x	x	√	√
Publish externally ELSST New Release: New Current and New Next	x	x	x	x	√
Create new users	x	x	x	x	√
Give permissions for roles	x	x	x	x	√
Manage tool configuration	x	x	x	x	√

UK Data Service

Example of SKOS concept

CRIMES AGAINST HUMANITY@en-GB

https://lod.data-archive.ac.uk/v2-skoselsst/en-GB/436aaebf-5594-4bbc-a7b8-07591ca3dd89

Property	Value
skos:altLabel	■ DELITOS HUMANITARIOS@es-ES ■ HUMANITAETSVERBRECHEN@de-DE ■ HUMANITARIAN CRIMES@en-GB ■ HUMANITÄTSVERBRECHEN@de-DE ■ NUSIKALTIMAI ŽMOGIŠKUMUI@lt-LT ■ RIKOS IHMISYYTTÄ VASTAAN@fi-FI
skos:broader	■ <https://lod.data-archive.ac.uk/v2-skoselsst/en-GB/5daecb52-714d-4e40-9be4-049de382f9c8>
skos:hiddenLabel	■ CRIMES AGAINST HUMANITY@en-GB
skos:inScheme	■ <https://lod.data-archive.ac.uk/v2-skoselsst/en-GB/>
dcterms:issued	■ 2017-05-15
skos:narrower	■ <https://lod.data-archive.ac.uk/v2-skoselsst/en-GB/b12a3cca-134c-4493-946a-cb284bb6bc54> ■ <https://lod.data-archive.ac.uk/v2-skoselsst/en-GB/f7d67282-82ea-4d98-9eac-c9fa2db8115d>
skos:prefLabel	■ BROTT MOT MÄNSKLIGHETEN@sv-SE ■ CRIME ÎMPOTRIVA UMANITĂȚII@ro-RO ■ CRIMENES DE LESA HUMANIDAD@es-ES ■ CRIMES AGAINST HUMANITY@en-GB ■ CRIMES CONTRE L'HUMANITÉ@fr-FR ■ DK_CRIMES AGAINST HUMANITY@da-DK ■ FORBRYTELSER MOT MENNESKEHETEN@no-NO ■ HUDODELSTVO ZOPER ČLOVEČNOST@sl-SL ■ NUSIKALTIMAI ŽMONIJAI@lt-LT ■ RIKOKSET IHMISYYTTÄ VASTAAN@fi-FI ■ VERBRECHEN GEGEN DIE MENSCHLICHKEIT@de-DE ■ ZLOČINY PROTI LIDSKOSTI@cs-CZ ■ ΕΓΚΛΗΜΑ ΚΑΤΑ ΤΗΣ ΑΝΘΡΩΠΟΤΗΤΑΣ@el-GR
skos:related	■ <https://lod.data-archive.ac.uk/v2-skoselsst/en-GB/7130642a-98a1-4ae6-9be0-00f178c4d128>
skos:scopeNote	■ OFFENCES CONSIDERED TO BE OF SUCH A MAGNITUDE THAT ARE ANSWERABLE TO THE WHOLE OF HUMANITY AND NOT TO A SPECIFIC JURISDICTION.@en-GB
rdf:type	■ skos:Concept

Download this resource as:
RDF/XML | Turtle

Requirements 2/6

- Workflow (behind the scenes)
Editorial workflow modalities (source language, target languages), validation of alphabetical, hierarchical and visual graph presentation

- System functions: management (behind the scenes)
PT management (create/change/delete/promote/demote), UF management, Hierarchies management, Notes management, versioning, import/export

UK Data Service

Requirements 3/6

- Validation and multilinguality (behind the scenes)
  Thesaurus creation/management, unique identifiers, prevention of term duplication, warning of near term equivalences, generated lists of allowed terms (based on rules and constraints), automatic symmetrical/reciprocal relationships, multilingual management Hub-Spoke model

- Search/browse/select (UI)
  search for PTs and UFs, autocomplete/wildcards, with/without language-specific diacritics, language-specific stemming, expanded search in all note types, filters, language-specific alphabetical glossary, switch between display modes

UK Data Service

Requirements 4/6

- **Results (UI)**
  Identify PTs/UFs, correct spelling (with diacritics), sorting options (Relevance, A-Z, Type), all translations identified by language abbreviations, mode options (expandable tree, listing, visual graph)

- **Concept views (UI)**
  Different modes of display, all information types (+/-), language equivalents, links to SKOS concept, version history, navigation options (hierarchy, multilingual, versions)

- **Visual graph (UI)**
  Spider network (UF, NT, BT, RT), switch language

UK Data Service

Requirements 5/6

- Edit function (behind the scenes)
  - SL editor, TL editor
  - linked/mapped thesauri (ELSST/HASSET)
  - two important functions re validation
    - automatically generated lists of allowed BTs, NTs and RTs
    - warning messages (prevent breaching rules)
  - actions are recorded and displayed in language-specific History pages in chronological order
  - actions are input for filtered search in History

UK Data Service

Requirements 6/6

- **History (UI)**
  Language-specific, concept/PT, action, BT/NT/RT/ Scope note etc., date, responsible, export options for reports (print, pfd, excel)

- **Suggestions (UI)**
  language-specific, suggest a change, comment on a suggestion, admin moderates and decides on status (accepted, rejected, under discussion), display options (Tree view), filters (= History, Edit pages), export options for reports (print, pdf, excel)

- **Bulk upload online, User guides, SKOS**

UK Data Service

Rules and constraints **mapping** HASSET – ELSST core terms 1/2

- a core term's BTs to the root also have to be core terms. All TTs that are not orphans must be core terms. TTs that are orphans, i.e. they do not have NTs, can be core or non-core.
[[PT = core] + [BTs = core]] in local tree and on every level
*[[PT = core] + [BTs = non-core]] in local tree and on every other level

- a core term's BTs to the root must be the same in both thesauri.
[[PT = core] + [BTs in A = BTs in B]] where A =Thesaurus A and B = Thesaurus B
*[[PT = core] + [BTs in A ≠ BTs in B]] where A =Thesaurus A and B = Thesaurus B

UK Data Service

Rules and constraints **mapping** HASSET – ELSST core terms 2/2

- a core term can have different NTs in Thesaurus A and Thesaurus B.
  [[[PT + NT1] in A] + [[PT + NT2] in B]]]

- a core term can have different RTs in Thesaurus A and Thesaurus B.
  [[[PT + RT1] in A] + [[PT + RT2] in B]]]

UK Data Service

Rules and constraints **no-mapping** HASSET – ELSST non-core terms

- a non-core term cannot be the same as all its NTs to its final node.
  *[[PT = non-core] + [PT = NT]] in local tree and on every other lower level

- a non-core term cannot be the same as all its RTs to its TT.
  *[[PT = non-core] + [PT = RT]] in local tree and on every other level

- none of a non-core term's RTs can be one of its BTs or NTs.
  *[[PT = non-core] + [[RT = BT] or [RT = NT]]]

- a non-core term cannot have a core NT.
  *[[PT = non-core] + [NT = core]]

UK Data Service

Editing 2 English **Monolingual** thesauri

- Process 1 English thesaurus
  Add new concept -> Edit concept -> Delete concept

- Management 2 English thesauri
  - Promote concept (to core)
    Demote concept (to non-core = delete in other thes.)
    – Restriction 1) Copy core PT/BT/NT/RT/SN
    　　　　　　　2) Not copy non-core NT/RT
  - Edit CORE terms: changes management
    – Restriction 1) Copy core TT/BT/NT/RT/SN(S)
    　　　　　　　2) Optional copy UN/HN

UK Data Service

Example of Edit concept

Editing SL **Multilingual** thesaurus

- Exclusive for SL editing/management

  - Adding new concepts

  - Relabelling existing concepts

  - Deleting concepts

  - Adding/deleting relationships: BTs, NTs, RTs

  - Adding/changing/deleting SNs and SNSs

- SKOS concept for all languages
  (currently under revision -> upgrade to SKOS-XL, XKOS;
  export + import)

UK Data Service

Editing TL spokes 1

- Process
  Not started -> Untranslatable OR Incomplete -> Complete

- Obligatory translation

  - Exact/near equivalence of PTs (+Untranslatable) which can be new or have been relabelled

  - Exact/near equivalence of Scope notes and Scope note sources

UK Data Service

Editing TL spokes 2

- Language-specific optional translation
    - Translating, adding, changing, deleting UFs
    - Translating, adding, changing, deleting Use notes
    - Translating, adding, changing, deleting History notes

- Language-specific no translations involved
    - Editorial note
    - Reindex: Y/N

Example Edit translation

SL and Language equivalents

TL and Language equivalents

Questions

Jeannine Beeken

jeannine.beeken@essex.ac.uk