

**In search of a survival kit for mappers:
Abstracting guidelines from complex mapping
examples. By Grete Seland, University of Oslo Library
EDUG mapping workshop – Napels, April 15th 2015**

O V E R V I E W	<p>Introduction (project, vocabularies and test mapping) What is challenging in mapping our vocabularies to Dewey? Why is mapping to a classification scheme so difficult? Which Dewey-specific issues should be approached? Conclusions (lessons learnt from test mapping, scenarios for application of mappings)</p>
1	<p>Humanities thesaurus: preferred term + UF, TT, BT, NT, RT, DEF</p> <p>Hearing impairment UF Hearing loss Hearing weakening Hearing defect</p> <p>TT Health BT Disability NT Deafness RT Hearing disorders Hearing impaired people DEF Designation of a state of complete or partial loss of ability to perceive or understand sounds</p>
2	<p>Science vocabulary: preferred term (single subject, or subject heading string) + UF</p> <p>Herbivores UF Plant eaters</p> <p>Algorithms Algorithms : History Algorithms : Popular science</p>

3	<p>Exact equivalence (exactMatch) – Humanities thesaurus</p> <p>Term equivalence between thesaurus term and caption: Public performances =EQ 791 (<i>caption</i>: Public performances) <i>Expressed in SKOS</i>: Public performances skos:exactMatch 791</p> <p>Conceptual equivalence, but not term equivalence: Koine Greek language =EQ 487.4 (<i>caption</i>: Koine (Hellenistic Greek))</p>
4	<p>Broader mapping (broadMatch) – Science vocabulary</p> <p>Aeromagnetic data BM 538.78 (<i>caption</i>: Magnetic surveys)</p> <p><i>Expressed in SKOS</i> Aeromagnetic data skos:broadMatch 538.78</p>
5	<p>Challenge 1: Concepts (“what does X mean?”)</p> <p>Humanities thesaurus concept: Cooperation (what is included/excluded as conceptual content?)</p> <p>Science voc.: Symmetri groups (requires topical knowledge)</p>
6	<p>Challenge 2: Mappings (where do I find this topic in Dewey? – “everywhere” vs. “nowhere”)</p> <p>Tobacco: “everywhere” - botany, ethics , religion , agriculture, human toxicology, production technology, customs, smuggling, etc.</p> <p>Specific terms: “nowhere” – symmetri groups, time-of-flight mass spectrometry, principal component analysis, reactive intermediates, etc. Frequent phenomenon in science voc.</p> <p>LCSH mappings might suggest relevant class number, e.g., “coronal mass ejections” at 523.75 (<i>caption</i>: Chromosphere and corona).</p>

7 Challenge 3: Relationships

Correspondence table for ISO and SKOS relationship types:

Types of mapping	Equivalence mappings		Hierarchical mappings		Associative mappings
ISO formal relationship types and symbols	exact equivalence =EQ	inexact equivalence ~EQ	broader mapping BM	narrower mapping NM	related mapping RM
SKOS	exactMatch	closeMatch	broadMatch	narrowMatch	relatedMatch

Easy choice of exact equivalence mapping (exactMatch):
Soul =EQ 128.1 (*caption: Soul*)

Should we also make a broader mapping (broadMatch)?
Soul BM 218 (*caption: Humankind, relative index: Soul--religion--philosophy of religion*)

Distinction between exact (=EQ) vs. inexact (~EQ) mapping:
Labor migration {relationship =EQ or ~EQ} 331.127
(*caption: Labor mobility*)

Topic in “class here”-note (i.e. “approximate the whole”):
inexact equivalence (closeMatch) as a rule?
Militia ~EQ 355.37 (*caption: Reserves, note: Class here home guards, home reserves, militia, ...*)

Topic in “including” note (i.e. “standing room”) – broad?
Deontological ethics BM 171.2 (*caption: Systems based on intuition, moral sense, reason; on duty and rights, note: Including deontology (nonconsequentialism), ...*)

Other examples of distinction between inexact mapping (CloseMatch) vs. broader mapping (broadMatch):
Feminist film theory {relationship ~EQ or BM} 791.43082
(*caption: Motion pictures + T1-082 Women*)

Broader mapping vs. associative mapping (relatedMatch):
Is shampoo a type of soap or another concept than soaps?
Shampoo {relationship BM or RM} 668.12 Soaps

	<p>Either several narrower mappings or one broader mapping: Collections (libraries) NM 026 (<i>caption</i>: Special libraries, <i>note</i>: Class here ... collections in specific subjects) Collections (libraries) NM 027 (<i>caption</i>: General libraries), or: one broader mapping? Collections (libraries) BM 020</p> <p>Should we make an associative mapping for the RT of the concept which we are mapping from the source vocabulary? Collections (libraries) RM 025.21 (<i>caption</i>: Collection development)</p> <p>If we make a mapping for each occurrence of “tobacco” in Dewey: Do we have to differentiate the relationships for each mapping? Several independent mappings vs. compounds.</p>
<p>8</p>	<p>Challenge 4: Context Cooperation {relationship BM or NM} 302.14 (<i>caption</i>: Social participation): Superordinate vs. subordinate context.</p> <p>Should we consider how the source concept is used in indexing? Potential for statistical mapping for specific vs. general terms.</p> <p>Example term from the science vocabulary: Vi improved No match in WebDewey. 5 records in catalogue with this index term, all classif. at 005.52 (<i>caption</i>: Word processing)</p> <p>Militia ~EQ 355.37 None of the bibliographic records contains this class number.</p>
<p>9</p>	<p>Comparing class numbers in mapping and bibliographic record. a) Match between mapping and classification</p> <p>A book on autism spectrum disorder: 4 index terms from the Humanities thesaurus: Autism, Diagnosis, Therapy and Psychotherapy. 6 Dewey mappings. 1 classification: 616.85882. Classification number matches the only Dewey mapping for Autism.</p>

10	<p>Comparing class numbers in mapping and bibliographic record b) Match between several mappings and built number</p> <p>A book on deaf people's citizen participation: 3 index terms from the Humanities thesaurus: Deaf people, Citizenship and Participation. 7 Dewey mappings (of which 1 is a table number). 3 classifications: 323.0420872 (citizen participation - deaf people), 305.9082 (deaf people, under sociology, groups of people) and 371.912 (students with hearing impairments)</p> <p>One of the classifications matches a combination of two of the mappings: Deaf people T1-0872 (<i>caption</i>: People with hearing impairments) Citizenship 323.042 (<i>caption</i>: Citizen participation)</p>
11	<p>Comparing class numbers in mapping and bibliographic record c) No match between mapping and classification</p> <p>A book on use of laptops in school teaching in Ethiopia: 6 index terms: Information technology, Learners, Ethiopia, Learning, Schools and Computers. 10 Dewey mappings (2 table no). None of the mappings matches the classification number assigned to the doc. in the bibliographic record.</p>
12	<p>Reality: A few facts about the Humanities thesaurus Non-typical as a thesaurus. Concepts with 2 or 3 BTs/TTs. Academic orientation (disciplines). 26 top-terms. Mixing of conceptual categories.</p> <ul style="list-style-type: none"> Archeology <ul style="list-style-type: none"> NT Geoarchaeology (generic – a type of archeology) NT Ancient monuments (not generic, but associative) <p>Qualifiers (parentheses) used for three different purposes:</p> <ul style="list-style-type: none"> - homonymy: Parasites (animals) vs. Parasites (people) - disciplinary context: Naivism (art) vs. Naivism (literature) - split compounds: Categories (linguistic) instead of Linguistic categories

13	<p>Technicalities: SKOS issues</p> <p>Handling of</p> <ul style="list-style-type: none"> - compound equivalence mappings - mappings to table numbers (auxiliary tables and add tables) - mapping to spans <p>Is SKOS the best option? What could be an alternative?</p>
14	<p>Which issues should the guidelines approach?</p> <p>Suggested topics for tomorrow's group discussions:</p> <p>Questions related to choices which need to be made according to ISO 25964-2 section 12.4 p. 31, concerning the second ISO point, i.e. "how much to differentiate the mappings"</p> <ul style="list-style-type: none"> • What would a pragmatic, applicable definition of equivalence between a thesaurus concept and a Dewey class have to state? • What are the criteria for distinctions between pairs of relationships, e.g.: <ul style="list-style-type: none"> - Exact vs. inexact equivalence (i.e. exactMatch vs. closeMatch) - CloseMatch vs. broadMatch (Comment: For topics which are not mentioned in a "class here" or "including" note, it is difficult to establish whether the thesaurus concept and the Dewey class are overlapping to some degree (and lead to closeMatch?), or whether the thesaurus concept as a whole is included in the Dewey class (in which case will still wonder: closeMatch or rather broadMatch?)) - CloseMatch vs. relatedMatch (degree of similarity within a category, or different concepts?) • When should we establish related mappings (relatedMatch): to implement the RT relationship in the source thesaurus (typically concepts from different categories), or to make a distinction between closeMatch and relatedMatch for subjects of the same category? Cf. ISO 25964-2 section 10 p. 25: "The dividing line between an associative mapping and inexact equivalence is ill-defined and subjective", which concerns degrees of similarity between concepts from the same category. <p>Questions related to Dewey-specific properties (pre-coordination, hierarchies, decimal structure) affecting the mapping procedure</p> <ul style="list-style-type: none"> • How are we to understand/define "concept" in Dewey as a target unit in mapping (cf. a Dewey class as a "container" for a group of objects – not a concept as a "unit of thought" as we are used to in thesauri)? • The understanding of broader/narrower in establishing mapping relationships – in what respect? In thesaurus terms (i.e. logically generic/partitive/instance relations), or in Dewey terms (super-/subordinate without logical relationships)? • Would it be useful to assign rules for the handling of "class here"

(closeMatch?) and “**including**” notes (broadMatch?), and how would an eventual application of closeMatch vs. broadMatch for this purpose “interfere” with the use of these relationships for topics which are not listed in such notes?

- What guidelines are needed to handle **mappings to table numbers** (both “numbered auxiliary tables”, as well as special “add tables” within main classes)?
- How could we apply the **context of the source vocabulary**
 - For the clarification of **conceptual content**?
 - Take into account the **disciplinary perspective**?
 - What about the **overall topical area** covered by the source thesaurus (e.g., “humanities” or “science”) – should this be considered in mapping? (Example: In mapping “materialism” to Dewey, we would arrive at class number 146.3 (materialism under philosophy). When this topic is used in a science vocabulary, it is in a science/mathematics meaning. Should we establish a mapping (broadMatch) to 510.1, even though “materialism” is not listed in this Dewey class (philosophy and theory under mathematics) - and disregard 146.3? Or rather, include both – considering that our overall goal is to establish a general thesaurus for which the entire vocabulary is mapped to Dewey?)
- Should **mapping to all disciplinary contexts in Dewey** be recommended as a rule? Irrespective of the kind of source vocabulary?
 - What about **thesauri, which are grouped according to a classification structure** (cf. ISO 25964-1 12.2.5.1 p. 76), as e.g. the humanities thesaurus at the University of Oslo Library? These kinds of thesauri, with a disciplinary orientation as top structure, is not treated in ISO 25964-2. In indexing and searching, these thesauri are treated as post-coordinated. However, the disciplinary structure adds an element of pre-coordination. If we were to consider disciplinary context in such a thesaurus in mapping to Dewey, it would be like mapping between two pre-coordinated vocabularies – virtually impossible (and probably the reason why LCSH has split their strings before mapping?). Should we just disregard the disciplinary context in the source vocabulary in mapping, i.e. only consider logical relationships, which provide help concerning conceptual content?
 - When mapping to all disciplinary contexts (e.g. tobacco), and a given topic occurs in various ways (caption, “class here note”, “including note”, table number, etc.): If we make several independent mappings, we would have to **differentiate between the relationship types in each mapping**. In a compound mapping, the compound would be provided with one mapping relationship for the compound. If we have to make several independent mappings for technical reasons – could the relative index be exploited in some way to save time/efforts?

- How could **sources external to the vocabularies** be used in the clarification of conceptual content (encyclopaedias, dictionaries, etc.)? What do we do when there is a discrepancy in the definitions provided in the vocabularies and external sources?
- How could **bibliographic data** be used - for the definition of conceptual content - and/or for statistical mapping (which might be especially relevant for specific terms)?

Questions related to technicalities and application of mappings in end-user tools

- Handling of relationship types in SKOS: **Technical restrictions due to SKOS**
 - How to handle **compound equivalence mappings** in SKOS, either way: a) “One-to-many” from thesaurus to Dewey can be handled as several independent mappings, b) There is no solution for “many-to-one” mappings from thesaurus to Dewey - e.g. women + leaders mapped to one class number for female leaders)
 - How to handle mappings to **table numbers** in SKOS - numbered auxiliary tables, as well as add tables.
 - How to handle mapping to **spans** (and what is the difference between “centred headings” and other types of number spans - in a WebDewey context?)
- When discussing technical solutions in SKOS: Do they **concern the actual mapping procedure, and/or the subsequent application** of mappings in end-user searching? Example: Compound mapping versus several independent mappings.
- Could the guidelines be more specific about **application of mappings in (pre)search tools**?
 - **Ranking order** of hits - e.g. exact, close, broad etc.?
 - Automatic **query expansion**?
 - **Exploitation of RTs** in the source vocabulary integrated in Dewey in an end-user search tool
 - How to use indexing tools (thesauri, Dewey) in end-user tools - for **presearching in several steps in the metadata** before being confronted with massive hits of retrieved documents.

15	<p>Scenarios for end-user tools: Utilizing metadata in several steps</p> <p>thesaurus term → docs (not applicable)</p> <p>thesaurus term → dewey class → docs (applic. for some specific topics)</p> <p>Dewey search with a rich entry vocabulary (synonyms + specific topics). Integrating RTs from the source vocabulary: A real contribution.</p> <p>How can we make better use of our indexing efforts in (pre)search tools?</p>
-----------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

“Problems worthy of attack prove their worth by hitting back.” (Piet Hein)