

Policy statement for the project Mapping to Norwegian WebDewey at the University of Oslo Library

The policy statement provides guidelines for the activities in the project at a general level. It expresses the aim and purpose for the project, as well as the principles and general conditions for how the project work will be conducted. Decisions made as part of the project should be based on the policy statement. Key concepts are defined in Section 4.

1 Aim: What we intend to do in the project Mapping to Norwegian WebDewey

Project aim:

The project aim is to map the vocabularies Humord ('humanities terms') and Realfagstermer ('science terms') to Norwegian WebDewey.

Subsidiary objectives:

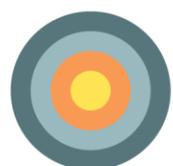
- 1) The project will contribute to the further development of a methodology for mapping to the Dewey Decimal Classification.
- 2) The mapping task will be performed as a computer-assisted intellectual process. The project will result in the development of a software tool for mapping of vocabularies, including the maintenance of mappings in conjunction with changes to the involved vocabularies.

2 Purpose: Rationale for the project

The purpose of the project Mapping to Norwegian WebDewey is to establish interoperability between vocabularies across languages and bibliographic databases. This will be achieved by a collective verbal access to the vocabularies involved, via Norwegian WebDewey.

The mapped vocabularies will make up the basis for improvements of end-user systems in the form of search facilities across collections and languages, as well as facilities for browsing of subject metadata before searching. Library staff will be offered support in indexing work in terms of connection points between selected index terms and corresponding classification, and vice versa.

The implementation of the project will include an improvement of the quality of the source vocabularies in terms of revisions and corrections made in conjunction with the mapping process. The project will contribute to an enrichment of Norwegian WebDewey in terms of an expansion of the number of verbal access points, due to the richer granularity in the subject areas specific to Humord and Realfagstermer. The project will also enrich Norwegian WebDewey with built numbers, as well as a quality assurance in the form of feedback to the editorial team for WebDewey concerning classes that require revisions.



The need for controlled vocabularies for Norwegian expert terminology extends beyond the library field. Mappings that are established in the project will be published as linked open data, available to other services on the Internet.

3 Background

The idea of mapping Norwegian vocabularies arose as an extension of a discussion about scenarios for the application of Norwegian WebDewey. This came in conjunction with a mapping seminar organized by the National Library, and Oslo and Akershus University College, in 2012. The fact that several vocabularies in English were mapped to the American WebDewey, gave rise to the idea of mapping Humord and Realfagstermer to Norwegian WebDewey. End users at the University of Oslo Library have to deal with several controlled vocabularies in conjunction with information searching, and they will benefit from a collective verbal access to the information resources.

In the period from March 2014 until March 2015, the University of Oslo Library conducted a pilot project on the development of a methodology for mapping of the Humord thesaurus to the Norwegian WebDewey. This constituted a separate activity within the larger project Towards a general Norwegian thesaurus. The project Mapping to Norwegian WebDewey elaborates on the pilot project, as well as two previous projects run by the Science Library at the University of Oslo. The first of these projects conducted a comparison of the terminology in Realfagstermer against the TEKORD thesaurus at the Norwegian University of Science and Technology. The second project explored methods for mapping between Realfagstermer and relevant sections of the preliminary Norwegian translation of the Dewey Decimal Classification.

The current project started in March 2015 and comprises three vocabularies: *Humord* is a thesaurus covering the humanities and social sciences with related disciplines. It is structured by hierarchical and associative relationships, and consists of terms representing approximately 18 500 concepts, as well as 8 500 synonyms for these. The Humord thesaurus is based on a collaboration between the universities of Oslo, Bergen and Tromsø, with the Humanities and Social Sciences Library at the University of Oslo as coordinator and secretariat. Humord is also used by The International Ibsen Bibliography and The Center for Studies of the Holocaust and Religious Minorities. *Realfagstermer* is a controlled, partially pre-coordinated subject vocabulary that mainly covers science, informatics, and mathematics. The vocabulary contains approximately 15 000 subject headings that can be used standalone or combined into strings, as well as approximately 2 000 synonyms for these. The vocabulary is multilingual (both of Norway's two official written languages, and in parts English). Realfagstermer is developed and maintained by the Science Library at the University of Oslo. Realfagstermer is also used by the University of Bergen Library, which is represented in the editorial team. *Norwegian WebDewey* is a complete translation of DDC 23 (Dewey Decimal Classification). It contains approximately 44 000 classes and is constantly expanded by built numbers. Norwegian WebDewey is only available online, and it is continuously updated.



4 Definitions

DDC: Abbreviation for Dewey Decimal Classification.

EDUG: Abbreviation for the interest group European DDC Users Group. EDUG aims at encouraging the development of Dewey Decimal Classification, in a collaboration between European user institutions and OCLC. EDUG's website is located at <http://edug.pansoft.de/tiki-index.php>.

ISO: Abbreviation for the International Organization for Standardization. In the current project, we use ISO 25964, especially Part 2 that concerns mapping, but also Part 1 that concerns thesauri.

Mapping: Mapping denotes a process of establishing relationships between concepts in two different controlled vocabularies. The mapping process starts in a source vocabulary from which corresponding concepts are sought in a target vocabulary. Mapping is also used to denote the result of mapping activity, i.e. a relationship between a concept in one vocabulary and one or more concepts in another vocabulary.

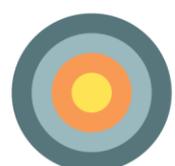
Mapping types and mapping relationships: According to the ISO standard for mapping (ISO 25964-2), there are three main types of mapping: Equivalence, hierarchical and associative. These mapping types give rise to five possible mapping relationships (with corresponding SKOS relationships provided in parenthesis): exact equivalence (SKOS: exactMatch, coded = EQ), inexact equivalence (SKOS: closeMatch, ~EQ), broader mapping (SKOS: broadMatch, BM), narrower mapping (SKOS: narrowMatch, NM), and related mapping (SKOS: relatedMatch, RM). Narrower mapping is never applied in mapping to Dewey Decimal Classification.

SKOS: Abbreviation for Simple Knowledge Organization System, an exchange format designed for the Semantic Web (W3C) for the representation of thesauri, classification schemes, and other types of controlled vocabularies.

Source vocabulary: A vocabulary that serves as a starting point when seeking a corresponding concept in another vocabulary (i.e. the target vocabulary). In the current project, the source vocabularies are made up of Humord and Realfagstermer.

Target vocabulary: A vocabulary in which a concept is sought corresponding to an existing concept in the vocabulary used as a starting point for the mapping (i.e. the source vocabulary). In the current project, the target vocabulary is Norwegian WebDewey.

Vocabulary/controlled vocabulary: A set of terms, subject headings, or notations that represent concepts and relationships between concepts, and that can be used in information retrieval.



5 Principles

5.1 Basic Documents

The policy statement is based on project funding applications, funding awards, and reports from the pilot project as well as the current project, both funded by the National Library. The stated principles are based on EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC), which again is based on the ISO standard for mapping (ISO 25964-2). The policy statement – as well as the basic documents on which it is based – will be used as a point of departure for the elaboration of the project's internal documents, e.g., mapping procedures, a collection of examples, and decisions related to which corrections should be made in the source vocabularies.

5.2 Basic assumptions

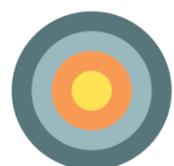
In line with EDUG's recommendations, the “hub structure” is used as the structural model, cf. the description of this model in ISO 25964-2 Section 6.4. In line with this model, Norwegian WebDewey makes up a central connection point – “a hub as the centre part of a wheel” – to which other vocabularies are mapped. We will use the four mapping relationships that are relevant in mapping to WebDewey, i.e. exact equivalence, inexact equivalence, broader mapping, and related mapping. The differentiation of mappings is based on an end-user perspective.

EDUG's recommendations provide an opportunity for a project to decide to map certain types of associative relationships. For example, if a project has made a decision to establish mappings between professionals and their disciplines, the source concept Physicians will be mapped to the class number for Medicine, in addition to the number for Physicians. The current project has not adopted mapping of such predefined types of associative relations. This project uses related mappings to establish mappings between associatively related concepts in individual cases. Related mappings are also used when we have doubts concerning the selection of a mapping relationship, but still want to establish a mapping. Furthermore, this project will not establish related mappings to represent associative relationships from within Humord and Realfagstermer, as these are internal matters concerning each of the source vocabularies.

The mappings will be coded in one direction in the mapping tool, from each of the source vocabularies to WebDewey. Exact equivalence is by definition bi-directional.

We will avoid establishing compound equivalence mappings because it is not possible to express them in the export format SKOS. In situations where a concept from the source vocabulary corresponds to a target concept distributed amongst several disciplinary contexts in WebDewey, we will establish several independent mappings to avoid compound equivalence. In the mapping process we will ignore cases where the source vocabulary has several independent concepts representing a conceptual content that is expressed as one complex concept in the target vocabulary (in practice, this affects general see also references in Humord).

The mapping process will undergo a quality control in the form of each mapping being inspected by two persons, one mapper and one reviewer. In case of doubt/professional disagreement, a third person should be consulted.



5.3 Scope

In the current project, we will map the source vocabularies universally to the target vocabulary, i.e. to all main classes in WebDewey. This is justified by indexing conventions for the source vocabularies at the University of Oslo Library, as well as with a view to a possible future general Norwegian thesaurus in which Humord and Realfagstermer will be integrated. For each source concept, we will map to all disciplinary contexts in WebDewey. The hierarchical context in Humord should be used for the clarification of conceptual content, but should not be used to restrict the hierarchical contexts upon which we are to establish mappings to WebDewey. The indexing conventions displayed in the bibliographic data will be used for supplementary clarification of concepts. We will establish mappings to schedule numbers, auxiliary table numbers (excluding T3), as well as built numbers. Mappings to optional numbers or discontinued classes will be avoided.

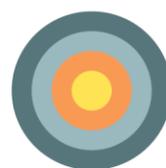
The project will map all concepts from the source vocabularies, including genre/form terms, geographical names, and subject heading strings (limited to strings of the type “content descriptive subject headings – content descriptive subject subheadings”, approximately 900 instances in Realfagstermer). When a concept is expressed as a schedule number as well as an auxiliary table number, we will establish mappings to both contexts. This applies for instance to linguistic concepts, and names for each language.

We will establish mappings to built class numbers when a source concept cannot otherwise be expressed in the target vocabulary, due to the richer granularity in the subject areas specific to Humord and Realfagstermer, as compared to WebDewey. Number building in conjunction with mapping is also in line with the convention of indexing as deeply as possible, as applied at the University of Oslo Library. Class numbers that are built in conjunction with the mapping process are to be registered in the WebDewey number building tool. For geographical names, the project will build numbers for history and geography as schedule numbers, as well as establishing mappings to the geography table using auxiliary Table 2. For ethnic groups, we will build numbers in the sociology table (at schedule number 305), as well as establishing mappings to auxiliary Table 5 for ethnic and national groups.

The project will establish guidelines for the treatment of non-typical features of Humord considered as a thesaurus, i.e. the handling of deviations from the standard for thesauri (ISO 25964-1). Structural clearing up of Humord (top terms, hierarchies) is not a part of the project, but will be directed to a possible future project for the development of a general Norwegian thesaurus.

5.4 ICT tools and technical clarifications

The mapping tool CCMapper will be developed by the technical staff members of the project, in collaboration with the German company Pansoft. The software is expected to handle the maintenance of mappings in conjunction with changes to the source or target vocabularies. Open source is a prerequisite. Completed mappings will be published as RDF/SKOS (without local extensions), as well as MARC 21. We differentiate between the technical solutions for data storage in CCMapper versus data exchange format.



The mapped vocabularies will be displayed in WebDewey in such a way that source concepts are represented by preferred terms. Mapping relationships should be specified together with the displayed mappings in WebDewey. Synonyms from the source vocabularies should be searchable and visible in WebDewey.

6 Resources

6.1 Economy

According to the funding award letter from the National Library, the project is funded with 2.086 million Norwegian crowns [equals 225 000 Euros] in 2015. For 2016, the project has received approximately 1.8 million Norwegian crowns [equals 195 000 Euros], including budget allocations under the contract with Pansoft. In addition to this, the University of Oslo Library contributes with its own share in the project, largely consisting of allocation of human resources to the project.

6.2 Staffing

Human resources for the mapping task will be brought into the project by allocation, as well as internal billing, of employees at the University of Oslo Library (librarians and academic librarians/subject specialists), as well as external appointments of personnel recruited from professional environments with expertise in Dewey classification and mapping. It may also be necessary to announce project positions for personnel that will attend to regular tasks at the University of Oslo Library, on behalf of internal employees who are working for the project.

Knowledge of Dewey Decimal Classification and thesauri are emphasized in preference to competence in the subject areas of the source vocabularies. The academic librarians/subject specialists at the University of Oslo Library will be consulted for conceptual clarifications. The project will conduct internal training in mapping, classification (including number building), and thesauri.

6.3 Time frame

The project runs two years counting from March 1st 2015. Allocated funds can be distributed over a longer period. The project will work according to a project milestones schedule, which will be enclosed with the policy statement.

6.4 Future resource requirements

When the vocabularies are mapped, a continuous maintenance of the mappings is mandatory because of changes to the source and target vocabularies. The project will establish guidelines for update management – with respect to content as well as technical solution. Maintenance of mappings will be carried out by the University of Oslo Library, which will log the changes to the source vocabularies. The National Library will announce WebDewey changes that have consequences for the mappings established by the University of Oslo Library.

The project will develop a plan for how the mapping expertise gained by the employees at the University of Oslo Library will be maintained after the project period has expired.



7 Further information

7.1 List of project members and contact information

Project Manager:

Unni Knutsen, Section Manager at the Humanities and Social Sciences Library,
Acquisition and Cataloguing Section

Email: unni.knutsen@ub.uio.no, Mobile: + 47-95777978

Project staff members - Mapping:

Berit Sonja Hougaard, Humanities and Social Sciences Library

Vibeke Stockinger Lundetræ, Humanities and Social Sciences Library s

Mari Lundevall, Science Library

Lembi Viola Kuldvere, Science Library

Grete Seland, Humanities and Social Sciences Library

Project staff members - ICT:

Dan Michael Heggø, Science Library

Are Dag Gulbrandsen, University Center for Information Technology

The project works in close collaboration with Elise Conradi at the National Library,
editor of the Norwegian WebDewey.

7.2 Project website

The project's website available from the University of Oslo Intranet at

<http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/>.

8 References

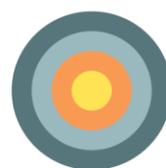
EDUG (2015 October 9th). *EDUG's Recommendations for best practice in mapping Involving Dewey Decimal Classification (DDC)*. Version 20151009. Retrieved from http://edug.pansoft.de/tiki-download_file.php?fileId=113

ISO 25964-1, i.e.: International Organization for Standardization. (2011). *Information and documentation: thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval*. Geneva: International Organization for Standardization.

ISO 25964-2, i.e.: International Organization for Standardization. (2013). *Information and documentation: thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Geneva: International Organization for Standardization.

Mapping to Norwegian WebDewey

Project application: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/prosjektsoknad.pdf>



Funding award letter: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/tildelingsbrev-mapping.pdf>

Status report and application for the second project year: http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/rapportering_nb_2015.pdf

Towards a general Norwegian thesaurus? Subproject *Methodology for mapping Humord to WebDewey*

Project application [scroll down to *På vei mot en generell norsk tesaurus*]: <http://www.nb.no/Bibliotekutvikling/Utviklingsmidler/Prosjektoversikter/Paagaende-prosjekt/Paagaar-t.o.m.-2015>

Funding award letter [requires login]: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/avsluttet/tesaurus-mapping/delte-dokumenter/tildelingsbrev-tesaurus.pdf>

Final report, Norwegian version: http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/metodikk_mapping_sluttrapport.pdf

Final report, English version: http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/methodology_mapping_final_report.pdf

