

**Traugott Koch<sup>1</sup>**

**NetLab, Lund University, Sweden**

**Technical Knowledge Center & Library of Denmark (DTV), Lyngby, Denmark**

**Heike Neuroth**

**State and University Library Goettingen (SUB), Germany**

**Michael Day**

**UK Office for Library and Information Networking (UKOLN), University of Bath, UK**

## **Renardus: Cross-browsing European subject gateways via a common classification system (DDC)**

**Abstract:** This paper presents the approach and first results of the classification mapping process in the EU project Renardus. The outcome in Renardus is a cross-browsing feature based on the Dewey Decimal Classification (DDC) and improved subject searching across distributed and heterogeneous European subject gateways. The paper presents the project's initial experiences and decisions, e.g. an investigation of the use of classification systems by Renardus partners' gateways, general mapping approaches and issues, the definition of mapping relationships and some information on technical solutions and the mapping tool. There is also a demonstration of the use of the mapping information in Renardus and the presentation of several features that have been implemented to aid end-user navigation in a large and deep browsing structure like the DDC. Classification mapping for cross-browsing is a labour intensive and complex effort which at the moment raises many open questions and leaves many more future potential work tasks than completed useful solutions.

### **1. The EU project Renardus**

Renardus<sup>2</sup> is a project funded by the European Commission as part of the Information Society Technologies (IST) programme, part of the European Union's 5th Framework Programme. Partners in Renardus include national libraries, research centres and subject gateway services from Denmark, Finland, Germany, the Netherlands, Sweden and the UK, co-ordinated by the National Library of the Netherlands. The project aims to develop a Web-based service<sup>3</sup> to enable searching and browsing across a range of distributed European-based information services designed for the academic and research communities - and in particular those services known as subject gateways. These gateways are services that provide access to Internet resources. They tend to be selective with regard to the resources they give access to, and are usually based on the manual creation of descriptive metadata. Services typically provide users with both search and browse facilities, and often offer hierarchical browse structures based on subject classification schemes (Koch & Day, 1997).

Predecessor projects like the EU project DESIRE<sup>4</sup> have already developed solutions for the description of individual resources and for automatic classification at the level of an individual subject gateway using established classification systems. Renardus intends to develop a service that can cross-search and cross-browse a number of distributed subject gateways through the use of a common metadata profile and by the mapping all locally-used classification schemes to a common scheme.

A thorough review of existing data models (Becker, *et al.*, 2000) was used as the basis for the agreement of a minimum set of Dublin Core-based metadata elements that could be utilised as a common data model. A comprehensive mapping effort from the individual gateways'

metadata element sets and content encoding schemes to the common profile has taken place. This provides the infrastructure for interoperability between all participating databases and thus is the necessary prerequisite for cross-searching

Enhanced subject access is considered to be one of the key services offered by subject gateways, and an important part of the Renardus service is its attempt to provide some kind of subject browsing across all participating gateways. The project has been, therefore, investigating ways that would enable users to browse a single subject hierarchy covering the content of all partner gateways. However, different gateway services use a wide range of classification schemes to provide browse access to Internet resources. These include well-known universal schemes, as well as a number of more subject specialised or locally produced systems. In order to accomplish consistent browse access to the content of Renardus partner gateways, all of the different classification systems in use need, therefore, to be mapped to a common classification system. The cross-browsing service in Renardus aims to mediate between the different classification systems in use by using the Dewey Decimal Classification (DDC) as a common switching language and browsing structure. An initial detailed description of the mapping effort and some preliminary guidelines are available from the project (Koch, Neuroth & Day, 2001a; 2001b).

The advantages of using classification systems to support subject access and topical navigation in large systems, i.e. interoperability, multilingual access, options to broaden or narrow searches, etc. are described elsewhere (e.g., Koch & Day, 1997). The Renardus service aims to give access to resources from all subjects, published world-wide and in many languages, and it is intended to be offered to an international multidisciplinary community of users. Taking these requirements into consideration, it appeared that utilising an existing universal classification system would be the most suitable tool upon which to build the common browsing structure in Renardus. Closer investigation demonstrated that DDC had important advantages, when compared with other schemes, for use in an application like the cross-browsing facility in Renardus. The main advantages lie in its online availability (e.g., WebDewey is a useful tool for the Renardus mapping process) and that its size and structure means that it is suitable for the task in hand. Other advantages include the scheme's global use, the large number of digital resources that have been classified with it and the speed and frequency of updates, especially with regard to the content of digital resources. Also advantageous are the research and methodological development efforts continually being undertaken by OCLC (Koch, Neuroth & Day, 2001a). The enhanced DDC also contains intellectually and statistically mapped vocabularies like the Library of Congress Subject Headings (LCSH) which are very useful for the Renardus mapping work.

The basis for the use of the DDC within the project is a research agreement with the scheme's owner, OCLC Forest Press. The license allows Renardus to use the full DDC classification system to construct and offer the Renardus cross-browsing pages. Co-operation with regard to methodological issues will also take place.

The classification and browsing solutions currently adopted by participating gateways are very heterogeneous. In order to prepare the mapping effort, it was necessary to conduct a thorough review of the schemes in use by partner gateways.<sup>5</sup> An analysis showed, for example, that many gateways use special subject schemes with a deep structure. For example, one gateway has 800 thematic classes structured in five levels that will have to be mapped. Other gateways' subject structures are not so extensive, with one or two levels of hierarchy and between 18 and 60 classes that will require mapping.

## **2. Mapping approach and issues**

A few practical principles are required to maintain consistency in the mappings and to ensure that the resulting Renardus browsing pages are balanced. Mapping relationships are

expressed between a pair of classes and not between a DDC class and individual resources. The mapping is carried out in one direction only, from the DDC to the local classification, i.e. the gateway's local browsing system. In order to help establish a balanced Renardus service at all times, it is suggested that gateways should finish mapping the top level of a local browse hierarchy completely, before moving progressively down through it. The ultimate goal is, naturally, to map all local classes to the DDC. The priority, however, is to map the most frequently used classes in the local gateway.

Many other issues will need to be discussed and solutions devised, possibly resulting in the periodic revision of the mapping guidelines. These issues may include specifics of how the DDC should be used to create a browse structure and how the mappings should be displayed in the Renardus browse interface. Other issues might include how deep the mapping should be on both sides (DDC and the local systems), how to treat local classes that contain both generalities and specialities, the exclusion of non-topical classes (e.g. auxiliary tables), the average number of allowed mappings, etc. Some of the subject areas that may provide the main focus of a gateway may be located deep within the DDC hierarchy. It is also not clear how the project should solve the conflict between the compact disciplinary structures that are often used in specialised subject classifications and the "shattering" of the same discipline within universal systems. For example, engineering is expressed in 800 classes within the specialised Ei classification system, but dispersed in about 2,300 categories in the DDC. Another problematic issue is the influence of the degree of subject overlap between the Renardus participants on the mapping practice. Similarly, Renardus has already discovered inconsistencies resulting from gateways' use of more than one classification scheme. For example, a subset of resources in one gateway might be classified using the DDC, while all resources are also classified with a different system; the one that would normally serve as the basis for the complete class-level mapping in Renardus. A permanent and very important issue is how to find the best trade-off between consistency, accuracy and usability in the Renardus cross-browsing service.

### **3. Mapping relationships**

Many other mapping projects, (e.g. those involved in conversions between two classification systems for use in OPACS or union catalogues), limit themselves to the establishment of simple connections between pairs of classes. They are often unspecific concerning defining the character and degree of the indicated equivalence. However, the structures and levels of detail, the vocabularies, languages and cultural contexts of the locally applied classification systems used by Renardus gateways and the DDC are very different. The project, therefore, assumes that a simple equivalence between the content of two classes will be rare. The same judgement has been made by other mapping projects, including CARMEN.<sup>6</sup>

In the Renardus subject browsing pages, users need to be advised that certain links from a DDC class, point to a class in a local gateway containing broader or narrower areas of content, or showing major or minor overlaps with the DDC class. This is especially true, as there will quite often be several mapping links to different classes found within a number of different gateways. One link might be fully equivalent, another might show just a minor overlap. The need for a more detailed specification of the degree of equivalence is even greater when the mapping between the local class and the DDC classes is used in the Renardus advanced search feature. The result list could be ranked according to the degree of relationship between the individual resource's local class and the DDC class used for searching.

Renardus has defined five distinct mapping relationships. The local class is deemed to be either fully equivalent, a narrower or broader equivalent, or has a major or minor overlap when compared with the DDC class. These relationships are influenced by the possible

relationships between sets in set theory and can be illustrated via Venn diagrams. This approach allows formal treatment and certain calculations on the relationships between the classes. "Fully equivalent" means that the subject content of the local page that one is linked to, is generally the same as the subject indicated on the Renardus browsing page. "Narrower equivalent" indicates that the subject content of the local page is a true subset of the browsing page, whereas "Broader equivalent" reflects the opposite scenario; the local page contains all of the subject content of the Renardus browsing page. Finally, "Major overlap" exists when the content of the local page represents a large part of the browsing page plus other related subjects. Conversely "Minor overlap" indicates some equivalence to part of the browsing page but that it may also include other related subjects. Renardus maps in one direction only, from the DDC to the local classification(s). The three types of equivalence require that one of the two classes is a true subset of the other, i.e. that it cannot also be mapped to another part of the classification scheme. Full equivalence is the intermediate situation where both classes are basically 100% equivalent. The two overlapping relationships require that parts of both classes clearly do not belong to the subject content of the other class. Thus certain logical rules apply which would permit a formal quality control of the mapping process.

existing relation (one-sided →)

GOK	relation	ddc	relevance	remark
ZU+000	LNE >	622		
cap 1	Mining: General			
cap 2	Mining and related operations			
url	http://www.Geo-Guide.de/cgi-bin/ssgfi/anzeige.pl?db=geo			

modify delete

Fig. 1: The Renardus mapping tool

#### 4. Technical solutions and tools

The main sources that are used for the classification mapping effort are the local classification systems and the enhanced DDC as presented in OCLC's CORC WebDewey. To support the practical effort, Renardus has adapted a mapping tool developed by the German CARMEN project. The Renardus mapping tool is Web-based and requires the open-source database software mySQL, an Apache Web server, JavaScript, and PHP scripts at the server side. The classification systems and mapping information are stored on different servers, partly for legal reasons. Each gateway participating in the mapping effort needs to provide a machine-readable version of the classification scheme (or schemes) that they use for use by the mapping tool. The user interface (Fig.1) consists of three main windows: one for the local

target classification, another for displaying and navigating the source classification (DDC). The third window receives and displays the mapping information, including relationships and notes. Mapping relationships are displayed as links in both classification windows. The tool has been adapted to create and store the mapping information in a MySQL database in a syntax specified by Renardus. This information is imported by Perl scripts into the main Renardus system in order to create the mapping links on the subject browsing pages and is also used by each gateway's local normalisation scripts in order to generate a DDC mapping for each resource in the local gateway's Renardus database.

The enhanced DDC is delivered by OCLC in several XML encoded data files with a XML DTD, tag/attribute information and additional information about hierarchy. They contain 25,500 main schedule entries (notations) and 35,700 different records. Using these files, an initial complete hierarchical set of web pages is generated allowing a user to navigate through the DDC structure. Completely empty branches in the lower part of the DDC hierarchy, however, can be removed from the display, assuming they are not required to assist as transitional steps during the browsing.

- [Top Level Subjects](#)
  - [Technology](#)
    - [Engineering](#)

## ▪ Mining and related operations ◀

- [Space mining](#)
- [Prospecting](#)
- [Excavation techniques](#)
- [\\*Mining for specific materials](#)
- [Nonextractive mining technologies](#)

[graphical navigation overview](#)

Find a different start-page for browsing:

### 'Mining and related operations' - Related Collections

*Fully equivalent:*

[Mining Engineering, General \(506\)](#) 

*Narrower equivalent:*

[Mining, General \(ZU 000\)](#) 


[Mining History \(ZU 400\)](#) 

Fig. 2: Renardus DDC browsing page for mining and related operations

## 5. The resulting cross-browsing feature in Renardus

The DDC mapping information is used in two different ways by the Renardus prototype, firstly to create the cross-browse service, but secondly to provide information for the advanced search feature. The aim of the cross-browse part of the Renardus is to allow users to navigate through the subject hierarchies of the DDC classification and to "jump" from a chosen class to related (i.e. mapped) classes and directories in the local subject gateways. This type of navigation can be called "browse and jump." The Renardus system specifies the different equivalences and degrees of overlap in the user interface. This approach allows the user to visualise the resources in the context of their local browsing structures and to continue browsing there (Fig. 2). The upper part of every page displays the available categories in the actual section of the hierarchy, with links to all levels above and one level below for users to follow. The lower half of the browsing pages shows one or more links to related resource collections. The local classification caption, the local classification code and the icon of the

gateway that the user would "jump" to when clicking on the link, are also displayed. The related collections are presented in a ranked order according to the recorded mapping relationship: fully equivalent classes are displayed first and minor overlapping classes last, thus to encourage the user to explore first the collections that are closest in coverage to the chosen DDC class.

Clearly, very large browsing structures, like that represented by the full DDC, need to provide additional assistance to guide users and features that supply an overview of options. Project investigations did not find any "tried-and-tested" solutions that Renardus could immediately apply. Therefore, the following preliminary navigation support features have been implemented for practical evaluation and criticism:

a) *Initial search for start page.* On all browsing pages (apart from the top level) a search box is offered to "Find a different start-page for browsing." Normally, several valid alternative browsing pages are displayed. This feature offers a short cut for users who know significant terms from a valid category elsewhere in the Renardus DDC structure. This may also be an option if users want to try their luck at finding a relevant entry for browsing, or if they have difficulty finding exactly where their main area of interest is hidden within the hierarchy.

From the alternative list, users can go to a selected browsing page or graphically explore the hierarchical environment of this subject for further navigation.

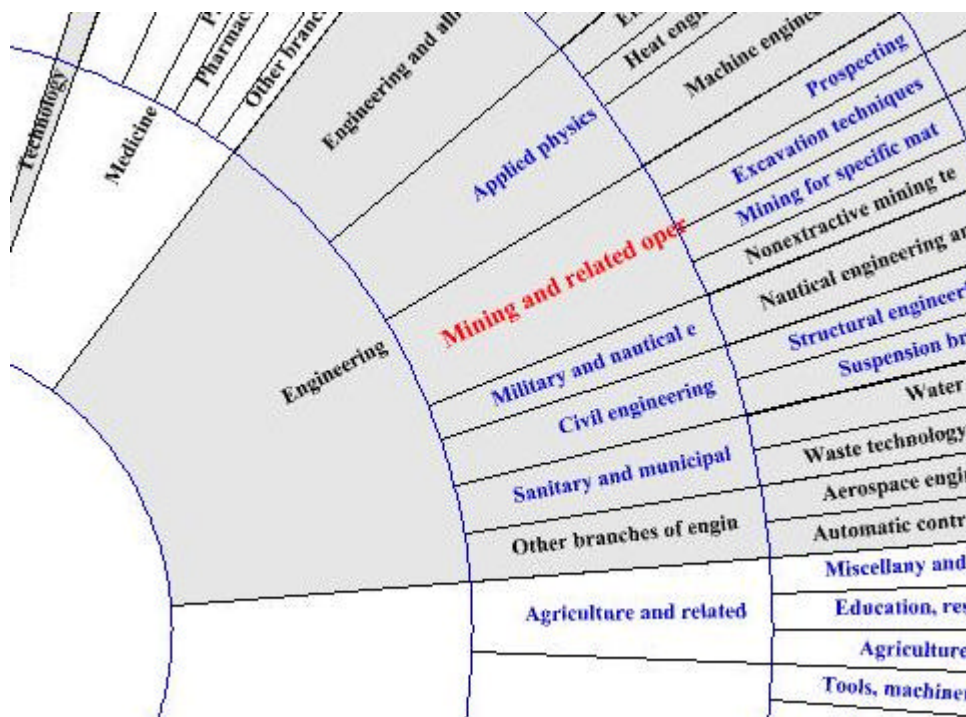


Fig. 3: Graphical navigation overview

b) *Graphical navigation overview.* A "Graphical navigation overview" (Fig. 3) is available on every browse page. It provides a visual overview of all the available categories that surround a chosen subject term, normally at one level above and two levels below within the hierarchy. Colours are used to help display the selected class within its context and all other classes that contain mappings. This feature is intended to increase the speed of users' navigation of the browse structure and to provide an immediate subject overview. Clicking on categories within the graphical display shows the relevant Renardus browsing page for this subject. An experimental text-based version of the browsing overview is also available.

c) *Merge.* Renardus also offers a short cut to viewing individual resource descriptions from all related collections with the feature: "Merge the resource-descriptions from all related

collections listed here." The main disadvantage with this "virtual browsing" is that users may lose some context and potential additional information if they do not explore the local gateway's browse structure. However, users may save time and will be shown an integrated list of resources from all related collections listed on the page, presented in the usual Renardus results display. The same kind of search can be carried out on the "advanced search" page by selecting the "DDC Classification" element for the search.

The DDC mapping information is also used in the Renardus advanced search feature. While the general subject element allows searching on all local subject information (e.g. uncontrolled keywords, controlled keywords from thesauri and subject headings, classification captions and notations, etc.) the "DDC classification" element enables searches to be made of the mapped DDC classes. A search normally opens an index scan window that allows the user to select from subject entries retrieved by the search. Layout and user interface solutions still need to be optimised based on user evaluation and usability research.

## **6. Future work**

The remaining mapping effort needs to be completed by Fall 2001, with the goal to map all of the local classes of participating gateways. Some loosely related gateways still have concerns about the amount of effort needed and the intellectual property rights involved. A more comprehensive quality control of the mapping is also required.

The project intends to formulate recommendations for subject access in gateways and broker systems (e.g., the choice of established systems, the granularity of the classification systems, continued local browsing features, etc.) and will give advice on the consistency, accuracy and usability of the mapping. It may also be possible to identify captions that may need to be adapted to reflect European usage, including experiments with Internet-adapted "end-user vocabularies" in co-operation with OCLC Forest Press. Furthermore, Renardus hopes to get additional agreement on the use of multilingual captions for the highest levels of the DDC in the cross-browse service. Another area of possible co-operation would be to investigate the usefulness of automatic mapping (and classification) experiments to support the mapping process.

The mapping methodologies and practical solutions developed by Renardus might need to be further refined. It would also be useful to explore the usability of the browsing service with real users. User interfaces for browsing in large distributed digital services are not as yet well developed. Co-operative development and standardisation efforts are needed in order to prepare vocabularies for distributed usage over the Internet with suitable encoding, identification and appropriate search protocols. In the longer term, the owners of established classification systems may need to become convinced of the need to provide mappings themselves and to maintain them as part of their vocabulary services, thus making the mapping task more sustainable.

The Renardus classification mapping effort is a large-scale experiment and, as yet, cannot build very much on theory or the practical experiences of others. It may raise more questions and problems than can provide solutions and answers. One has to hope that the effort can be continued, embedded in broad international co-operation, as a combined development and research effort that can influence future solutions.

## **Notes**

1. Part of the preparatory effort of T. Koch was supported by a Visiting Scientist scholarship at the Maastricht McLuhan Institute, European Centre for Digital Culture, Knowledge Organisation and Learning Technology, The Netherlands in August-September 2000.
2. The Renardus project Web-site can be found at: <http://www.renardus.org/>

3. A beta version of the Renardus broker pilot can be viewed at: <http://renardus.lub.lu.se/>
4. DESIRE II (<http://www.desire.org>): automatic classification efforts:  
<http://www.lub.lu.se/desire/DESIRE36a-overview.html>
5. Renardus project overview of classification systems in use - follow the link: "Overview Classification Systems": <http://renardus.sub.uni-goettingen.de/wp7/d7.4/>
6. CARMEN (Content Analysis, Retrieval and MetaData: Effective Networking) project:  
AP12: <http://www.bibliothek.uni-regensburg.de/projects/carmen12/>
7. See the illustration at:  
[http://renardus.sub.uni-goettingen.de/conferences/elag2001/mapping\\_language.gif](http://renardus.sub.uni-goettingen.de/conferences/elag2001/mapping_language.gif)

## References

- Becker, H. J., *et al.* (2000). *Evaluation of existing data models*, Renardus D6.1.  
<http://www.renardus.org/deliverables/>
- Koch, T. and Day, M. (1997). *The role of classification schemes in Internet resource description and discovery*, DESIRE D3.2 (3).  
<http://www.ukoln.ac.uk/metadata/desire/classification/>
- Koch, T., Neuroth, H. and Day, M. (2001a). *DDC mapping report*, Renardus D7.4.  
<http://renardus.sub.uni-goettingen.de/wp7/d7.4/>
- Koch, T., Neuroth, H. and Day, M. (2001b). *DDC mapping guidelines*, Renardus D7.4 (internal deliverable).