Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

Maren Brodersen / Lavinia Hodoroaba

Navigation in the "KOBV-Informationsportal"

- Conceptual View -

Gefördert
von der Senatsverwaltung für Wissenschaft, Forschung und Kultur des Landes Berlin,
vom Ministerium für Wissenschaft, Forschung und Kultur des Landes Brandenburg
und von den Mitgliedsbibliotheken des KOBV

Navigation in the "KOBV-Informationsportal"

- Conceptual View -

Maren Brodersen / Lavinia Hodoroaba

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

ZIB-Report 02-54

Dezember 2002

Abstract

The KOBV-Informationsportal aims to be a universal gateway to the sources of information hosted by the partner libraries from the Berlin-Brandenburg area. Due to the large number of these sources, an intuitive navigation is an essential component of the portal. The navigation-component should preserve the partner libraries' independence and overcome their administrative and technical differences.

This paper proposes a collection-level navigation with four dimensions: the sources' subject areas (e.g. the first two levels of DDC), the sources' type (e.g. e-journals, databases, OPACs, etc.), the sources' location (e.g. Berlin, Brandenburg) / the library that hosts that source and the sources' accessing state (e.g. free, restricted, etc.).

Keywords: Portale, KOBV, Navigation, Collection-level

CR: H.4

Contents

I. Introduction 3
II. The Navigation Concept in the KOBV Informationsportal 3
II.1. The Dimensions of the Navigation Structure 3
II.2. How to Omit the Navigation Structure 4
II.3. The Visual Appearance of the Navigation Structure 4
II.4. The Graphical Navigation 5
II.5. Populating the Navigation Structure 6
II.6. The Portlet Approach 6
III. General Considerations regarding Navigation 7
III.1. The Subject Area Dimension 7
III.2. The Type Dimension 13
III.3. The Location / Library Dimension 14
III.5. The Accessing State Dimension 14
IV. The Collection Concept 14
IV.1. The Motivation 14
IV.2. The Definition 15
IV.3. Examples of Collections 16
IV.4. The Granularity of a Collection 16
V. Possible Implementations 17
V.1. The KOBV Collection Metadata Model 17
V.2. The Collection Description Service 20
V.3. The Collection Discovery Service 21
V.4. The Collection Customization Service 22
V.5. Pros and Cons of the Navigation at the Collection Level 22
V.6. Possible Improvements 23
VI. Conclusions 24
References 25

I. Introduction

The KOBV Informationsportal has entered the second stage of its evolution and intends to provide integrated information services for the Berlin-Brandenburg area. To the KOBV users should be offered a single point of access to universal-interesting physical and electronic **sources of information** such as OPACs, collections of links, A&I databases, subject gateways, e-journal systems, etc. The word **resource** could be used synonymous with a source of information.

The ways of attaining the goal of the KOBV Informationsportal are influenced by the KOBV decentralization policy, the resources' heterogeneity from the area, and the KOBV Informationsportal's target group. In other words, the implementation of the KOBV Informationsportal should be performed preserving the partner libraries' independence, overcoming their **administrative and technical differences** and satisfying the target group's needs. Due to its goal i.e. to be an universal gateway for the Berlin-Brandenburg area, the target group of the KOBV Informationsportal contains experienced users, who know where to look for the information in which they are interested in, as well as novice users, who have little or no knowledge at all about which are the possible resources of interest for them.

Taking into account the large number of resources offered by the portal, both user categories require a complementary option to search for retrieving information, **a navigation option** in the portal's information environment. But, whilst for the experienced users, this navigation option could be materialized just in an information about the available resources, novice users need more guidance in discovering resources of interest. The novice users need to navigate **by browsing** categories of resources. The finer the granularity of the **navigation/browsing structure**, the better guidance for the novice users.

On the other hand, the navigation begins when the user interacts with the portal's information environment and could finish when the user discovers the resource(s) that he has looked for. The number of steps the user should perform in order to discover the resource(s) of interest could be selection criteria in using a portal. In summary, the granularity of the portal's navigation structure should be large enough – in order to guide the novice users, and in the same time should be small enough – in order not to waste the time of the advanced users. Also about the navigation structure's design, it is worth to emphasize that its visual appearance should be self explanatory, consistent, not overwhelming and in compliance with the W3C Recommendations for the Web Content Accessibility[31].

II. The Navigation Concept in the KOBV Informationsportal

II.1. The Dimensions of the Navigation Structure

In the KOBV Informationsportal, the users should be guided in the information environment based on the **resources' subject areas** (e.g. social science, biology, etc.), **resources' type** (e-journals, databases, OPACs, etc.), **resources' location** (Berlin, Brandenburg, etc.) / the library that owns the resource and the **resources' accessing state** (e.g. free, restricted, etc.). In other words, in finding resources of interest, the users should be able to browse categories of resources based on their covered subject area, type, physical location / library and accessing state. These four ways of navigating to a resource will be called, from now on,

dimensions of the navigation/browsing structure. The subject dimension is a two level dimension for allowing a more in-depth research on a specific topic, while the other three are one level dimensions. The possible values of these dimensions will be called, from now on, **options**.

Taking into account the size of the navigation structure (four dimensions), this could be further divided in **primary** and **secondary navigation**. The secondary navigation dimensions will allow the user to navigate just within the specified range of resources set up by choosing an option from the primary navigation structure. Which dimension(s) is/are part of the primary navigation structure and which of the secondary one, could be established based on the answer to the question: what dimension(s) will be most probable chosen when the KOBV user interacts with the KOBV Informationsportal? Taking into account that the time spent in discovering a resource of interest in a portal is an important selection criterion in using that portal, the KOBV user's time can be improved, for example, by providing at a first glance the number of the resources related to a subject area grouped by their types. In other words, the portal could provide as **dimensions of the primary navigation** structure the resource's **subject area** and its **type**. In consequence, the other two dimensions, namely the resource's **location** / **the library** and the resource's **accessing state**, could be part of the **secondary navigation**.

Speaking about the primary navigation, based on behavioral patterns, could be stated that a user is more interested in finding all the resources in a certain subject area, than in the available types. In consequence, in the **primary navigation** framework a distinction should be made between the **primary dimension** – the resource's **subject area** and the **secondary dimension** – the resource's **type.** The secondary dimension will list only the resources selected by the primary dimension. In the **secondary navigation** both dimensions could be perceived as of the same importance for the user

II.2. How to Omit the Navigation Structure

To note that the design of this navigation structure could be perceived as non-compliant with the requirement of keeping the balance between the novice users' needs and experienced users' needs. It is true that the design favors novice users, but a portal should also offer a customization service of the user's information environment. In the KOBV Informationsportal, this service could contain an additional feature which allows the user to specify what view he would prefer to receive when he logs-in: the navigation structure or directly the view with his selected resources. With this feature, the advanced users could by-pass the complexity of the navigation structure.

II.3. The Visual Appearance of the Navigation Structure

About the visual appearance of the browsing structure, it should be emphasized that the four dimensions of the navigation structure could be implemented using different styles, e.g. navigation buttons, navigation bars, links, image map, animated graphics, drop-down menus, drop-down lists, etc. Taking into account the requirements for the visual appearance of the browsing structure emphasized in Chapter I, the decision which of these styles should be used is based on the fact that the user should have as much as possible from the navigation structure at a glance, he should be able to use the browsing structure without the need, firstly, to learn it, and once a style has been used in a specific context, its significance should be preserved.

A possible combination of styles for the KOBV Informationsportal's browsing structure consist of using **links for the primary navigation** and **drop-down lists for the secondary navigation**. A prototype of this visual appearance could be reached at http://se3.kobv.de:2727/cgi-bin/navigation. The prototype is implemented on Solaris platform, using open source software (MySQL, CGI::XMLApplication, XSLT).

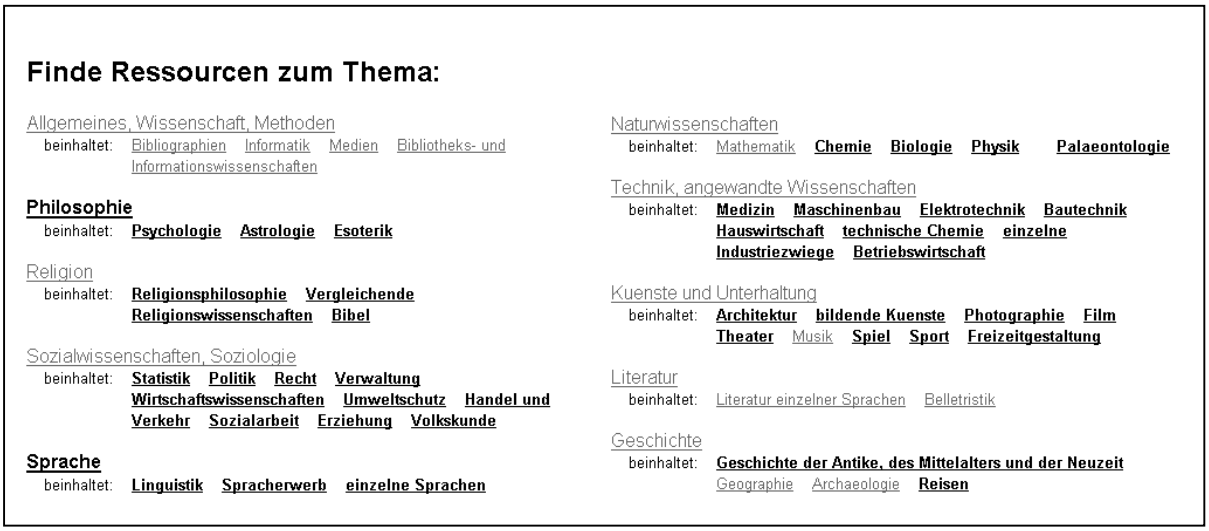

Fig.1. The visualization of the subject area dimension

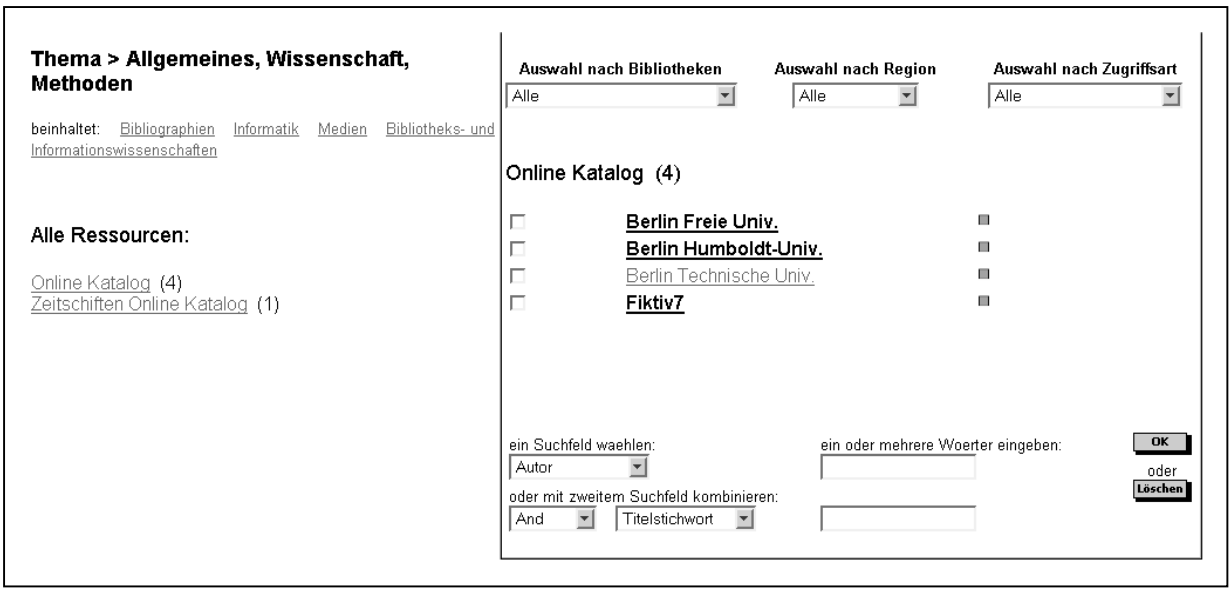

Fig. 2. The visualization of the type, library/location and accessing state dimensions

II.4. The Graphical Navigation

The KOBV Informationsportal could also provide a **graphical navigation** of these four dimensions. But because of the technical requirements of a graphical navigation (usually, Internet Explorer 4 or higher, Netscape 4 or higher and a Java support), this could be seen just as an add-on and not as the only available solution.

About the graphical navigation, at this time, it is emphasized that it could be one of the products available on the market, with the necessary customization for the KOBV Informationsportal, and should provide the navigation over the all four dimensions.

II.5. Populating the Navigation Structure

No matter which kind of browsing the user follows, graphical or textual, to the user will be presented, according to the selected criteria, the list of resources ordered alphabetically, for example. For each resource, the name, a short description, the URL (for online resources), a link to a long description and a graphical sign for the resource's accessing state will be listed. If a resource's accessing state is free or restricted, then the graphical sign could be a green circle for a free accessible resource and a red circle for the restricted one. The explanation what the user should do in order to gain access to the resource, could be part of the long description section of that resource. It is worth to point out here that a combination of browse and search could help the user in discovering the relevant resources in the smallest number of performed steps. For this reason, for the resources that are searchable, a check-box could appear for allowing the user to select certain resources for searching. Also, the terms for searching could appear on the screen – for example, at the bottom of the screen. For all online resources, the user could follow the URL in order to access the original interface of that resource. Having in mind that a graphical sign could be faster grasped than a text, some dimensions (or all dimensions) of the navigation structure could also incorporate a graphical representation.

II.6. The Portlet[1] Approach

Usually, after performing a search, if the user wants to browse further, he should use the back button or follow a link in order to access the browsing screen again. Because of the large number of the records available for searching in the KOBV portal, the number of hits for a search could be too large to fit into a screen and the user might scroll in order to see the results.

By using a portlet approach, the user can have all in-one-sight. The KOBV Informationsportal could contain one portlet for each service: one for browsing, one for searching, one for the news channel and so on. Using portlet technologies, to the user could be presented the browsing structure and the results of searching in one screen, without being too overwhelming for the user and without constraining the user to scroll. The browsing structure could appear in a portlet, for example in the left one, while the results of the cross-search or the resource chosen for searching with its original interface could appear in another portlet, for example in the right one. This could also be a good strategy in helping the user not to get lost when he experiences another resource's interface and to provide a means for coming back to the KOBV Informationsportal.

[1] "Portlets are the visible active components end users see within their portal pages. Similar to a window in a PC desktop, each portlet owns a portion of the browser … where it displays results." – WebSphere Portal; http://www-3.ibm.com/software/webservers/portal/portlet.html

At this time (Jan. 2003), there are products that support the portlet technology in the public domain (e.g. the Apache Jetspeed project [1] as well as in the commercial domain (e.g. the Oracle Application Server Portal [2] and the IBM WebSphere Portal Server [3]).

III. General Considerations regarding Navigation

In order to turn this vision into reality, **general issues** related to constructing a navigation structure in a distributed and heterogeneous environment, and **specific issues** related to the KOBV environment should be addressed. It is well-known that it is difficult to implement flexible "browsing" interfaces targeted at multiple partners. The main difficulties come from the differences in the values of the navigation structure's dimensions and in the indexing practices – which vary based on local needs. The values used by partners for the navigation structure's dimensions are based on different controlled vocabularies (classification schema, thesauri, etc.) and/or controlled values (e.g. values from ISO 639-1 two-character language code, ISO 639-2 three-character language code, ISO 3166 (2-letter-code) for countries, ISO 8601 (YYYY-MM-DD) for dates, etc.); standard ones or homegrown. In broader terms, for building a navigation structure, firstly, there is a need to overcome the lack of a unified controlled metadata and agree on a semantic for each dimension of the browsing tree and secondly, to perform a mapping between these dimensions and the corresponding dimension of the partner's resource.

In order to figure out the specificity of the KOBV environment, a study has been undertaken on the resources hosted by the libraries from the Berlin-Brandenburg area concerning the dimensions of the browsing structure.

III.1. The Subject Area Dimension

The most debated dimension is the **resource's subject area**. Usually, the subject area dimension is implemented based on a classification schema. Terms from thesauri and classes from classification schemas are used in browsing as means for improving subject access. At the resource level, from the undertaken study could be stated that there are resources for which no classification schema is used, but there exists a thesauri in use. There are also resources for which a classification schema and a thesauri are used (e.g. at the Hochschule für Film und Fernsehen "Konrad Wolf"). At the partner library level could be observed that a library can use more than one classification schema for its resources, or can use classification schemas and thesauri or just thesauri. Some of the used classification schemas are homegrown classification schemas (e.g. Senatsbibliothek or Technische Fachhochschule). The same observation is valid for the thesauri (e.g. Alice-Salomon Fachhochschule für Sozialarbeit und Sozialpädagogik or Ibero-Amerikanisches Institut). The first question that naturally arises is: which classification schema should be used for the browsing structure? Then arise implementation questions related to ways of overcoming these differences in the categorization of resources. For the implementation questions, possible answers will be outlined in the chapters IV and V.

In order to answer the first question, the trade-off between the portal's interoperability with currently projects in the area and the specificity of the KOBV and the libraries from the Berlin-Brandenburg area, should be taken into account. It is worth to be noted that projects such as RENARDUS[4], the German Virtual Library[5] use DDC[6] for the browsing structure, the

SSG-FI project uses GOK[7] as the primary subject classification, DDC[6] as the secondary subject classification and the BK[8] as the tertiary subject classification. In consequence, for the interoperability goal, the use of DDC for the navigation structure could be the right decision.

On the other hand, in the KOBV Suchmaschine[9], the BK[8] is used on the Datenbanken page for grouping the partner libraries OPACs. The same classification schema, namely BK[8], is also used in the Bibliothekenführer[10]. From the study could be concluded that not one of the KOBV partner libraries use BK[8] for the classification of their resources, five libraries use RVK[30] and two libraries use a classification based on GHB. Taking into account the specificity of the KOBV and the libraries from the Berlin-Brandenburg area, the BK[8] or RVK[30] could be the right decision. But due to the large number of categories on the first level (51 categories in comparison with DDC[6] that has 10), BK does not meet the design goal of a navigation structure. A correspondence between the classes of these three most probable used classification schemas is presented below:

DDC	BK	RVK
000 Allgemeines, Wissenschaft, Methoden beinhaltet: 004 Informatik 010 Bibliographie 020 Bibliotheks- und Informationswissenschaft 030 Enzyklopädien 050 Zeitschriften und Serien 060 Allgemeine Organisationen und Museen 070 Medien, Journalismus, Verlagswesen 090 Handschriften, seltene Bücher	▪ Allgemeines (Nachschlagewerke, Bibliographien) ▪ Information und Dokumentation ▪ Kommunikationswissenschaften (Presse, Hörfunk, Fernsehen) ▪ Wissenschaft und Kultur allgemein ▪ Informatik ▪ Geisteswissenschaften allgemein	▪ Allgemeines ▪ Informatik
100 Philosophie beinhaltet: 130 Parapsychologie, Astrologie, Esoterik 150 Psychologie	▪ Philosophie ▪ Psychologie	▪ Philosophie ▪ Psychologie

DDC	BK	RVK
200 Religion beinhaltet: 210 Religionsphilosophie, Religionstheorie 220 Bibel 230 Christliche Theologie 290 Vergleichende Religionswissenschaft, nichtchristliche Religionen	▪ Theologie, Religionswissenschaften	▪ Theologie, Religionswissenschaften
300 Sozialwissenschaften, Soziologie beinhaltet: 310 Statistik 320 Politik 330 Wirtschaftswissenschaften 333 Umweltschutz 340 Recht 350 Verwaltung 355 Militär 360 Sozialarbeit 370 Erziehung 380 Handel, Kommunikation, Verkehr 390 Volkskunde	▪ Sozialwissenschaften allgemein ▪ Sozialpädagogik , Sozialarbeit ▪ Soziologie ▪ Pädagogik ▪ Bildungswesen ▪ Politologie ▪ Ethnologie, Volkskunde ▪ Recht ▪ Verwaltungslehre ▪ Verkehrswesen, Verkehrstechnik ▪ Volkswirtschaft (inkl. Wirtschaftswissenschaften) ▪ Umweltforschung, Umweltschutz ▪ Geographie, Raumordnung, Städtebau	▪ Soziologie ▪ Pädagogik ▪ Politologie ▪ Rechtswissenschaft ▪ Ethnologie (Volks- und Völkerkunde) ▪ Wirtschaftswissenschaften

DDC	BK	RVK
400 Sprache beinhaltet: 410 Linguistik 420 Englisch 430 Deutsch 439 Andere germanische Sprachen 440 Romanische Sprachen, Französisch 450 Italienisch 460 Spanisch, Portugiesisch 470 Latein 480 Griechisch 490 Übrige Sprachen 491.7-.9 Slawische und baltische Sprachen	▪ Sprach- und Literaturwissenschaft ▪ einzelne Sprachen und Literaturen	▪ Allgemeine und vergleichende Sprach- u. Literaturwissenschaft ▪ Klassische Philologie
500 Naturwissenschaften beinhaltet: 510 Mathematik 520 Astronomie 530 Physik 540 Chemie, Kristallographie 550 Geowissenschaften 560 Paläontologie 570 Biowissenschaften 580 Pflanzen 590 Tiere	▪ Mathematik ▪ Physik ▪ Chemie ▪ Biologie ▪ Geowissenschaften ▪ Astronomie ▪ Naturwissenschaften allgemein	▪ Mathematik ▪ Allgemeine Naturwissenschaften ▪ Geologie u. Paläontologie ▪ Physik ▪ Chemie u. Pharmazie ▪ Biologie

DDC	BK	RVK
600 Technik, angewandte Wissenschaften beinhaltet: 610 Medizin 620 Ingenieurwissenschaften 630 Landwirtschaft 640 Hauswirtschaft 650 Betriebswirtschaft 660 Technische Chemie 670 Industrielle Fertigung, einzelne Industriezweige 690 Bautechnik	▪ Technik allgemein ▪ Bauwesen (inkl. Architektur) ▪ Medizin (inkl. Pharmazie) ▪ Bergbau ▪ Chemische Technik, Umwelttechnik, verschiedene Technologien ▪ Maschinenbau, Energietechnik, Fertigungstechnik ▪ Werkstoffkunde ▪ Elektrotechnik ▪ Hauswirtschaft ▪ Arbeit, Handwerk, Dienstleistungs-gewerbe ▪ Land- und Forstwirtschaft ▪ Tiermedizin ▪ Betriebswirtschaft	▪ Technik ▪ Medizin ▪ Land- u. Forstwirtschaft, Gartenbau, Fischereiwirtschaft, Hauswirtschaft
700 Künste und Unterhaltung beinhaltet: 710 Landschaftsgestaltung, Raumordnung 720 Architektur 730 Bildende Kunst, Kunsthandwerk 741.5 Comics, Cartoons, Karikaturen 760 Drucken, Vervielfältigung 770 Photographie 780 Musik 790 Darstellende Kunst, Freizeitgestaltung 791 Hörfunk, Fernsehen, Film 792 Theater, Tanz 795 Spiel 796 Sport	▪ Kunstwissenschaften (inkl. Kunst-geschichte) ▪ Einzelne Kunstformen ▪ Bauwesen (inkl. Architektur) ▪ Theater, Film, Musik ▪ Sport, Freizeit, Erholung	▪ Kunstgeschichte ▪ Musikwissenschaft ▪ Sport

DDC	BK	RVK
800 Literatur beinhaltet: 810 Literatur der USA 820 Englische Literatur 830 Deutsche Literatur 839 Literatur der übrigen germanischen Sprachen 840 Französische Literatur 850 Italienische Literatur 860 Spanische und portugiesische Literatur 870 Lateinische Literatur 880 Griechische Literatur 890 Literatur der übrigen Sprachen 891.7-.9 Slawische und baltische Literatur Belletristik	▪ Sprach- und Literaturwissenschaft ▪ Einzelne Sprachen und Literaturen ▪ Belletristik	▪ Germanistik, Niederlandistik, Skandinavistik ▪ Anglistik, Amerikanistik ▪ Romanistik ▪ Slavistik ▪ Klassische Philologie
900 Geschichte beinhaltet: 910 Geographie, Reisen 920 Biographie, Genealogie 930 Alte Geschichte, Archäologie 940 Geschichte Europas 943 Deutsche Geschichte 950 Geschichte Asiens 960 Geschichte Afrikas 970 Geschichte Nordamerikas 980 Geschichte Südamerikas 990 Geschichte der übrigen Welt	▪ Geographie, Raumordnung, Städtebau ▪ Geschichte (inkl. Archäologie)	▪ Geschichte ▪ Klassische Archäologie ▪ Geographie

For the browsing structure of the KOBV Informationsportal, the use of a homegrown classification schema is a barrier in the way of attaining the interoperability goal. In summary, for the subject area dimension the use of the first two levels of DDC could be DDC[6] a solution. But it is also possible that the subject area dimension be based on many classification schemas, in which case it should be presented on the portal's information environment as a drop-down list. The options listed for the navigational button should be dynamically created based on which classification schema is selected from the drop-down list. But this implementation will add more complexity to an already complex navigation structure and could be generate performance problems.

III.2. The Type Dimension

The **resources' type** is the second dimension of the navigation structure and the possible options of this dimension are based on the results of the undertaken study. For this dimension there exits an agreed vocabulary, namely DCMI Type Vocabulary[28], but it is not used by the partner libraries. More than this, the DCMI types do not cover all the specific types encountered in the KOBV environment. It is worth to be emphasized here that some OPACs contain only metadata about books while others contain metadata about books and journals, for example. This difference in the content of an OPAC should be presented to the KOBV user during the navigation process in order to enable him to discover the exact type of material in which he is interested in.

A proposed list of these categories and their abbreviations is presented below:

Category	Abbreviation
Katalog/OPAC – allgemein, umfaßt Bücher und/oder Zeitschriften etc.	OPAC
Zeitschriften-OPAC	OPAC-Z
Elektronische Zeitschriften / E-Journals	COLL-EZ
Datenbanken	DB
Rezensionen	REZ
Index & Abstracts	IA
Forschungs- und Tagungsberichte	FORSCH
Preprints/Arbeitspapiere/Graue Literatur	PREP
Hochschulschriften (Diss., Habil., Magister, Diplom)	HOCHSCH
Linksammlungen	LINKSAM
Bibliographien	BIB
Videos	VIDEO
Audios	AUDIO
Software	SOFTW
Bilder / Images	IMAGE
Karten	MAPS
Bibliotheken	BIBL
Lehrmaterial/Tutorials	TUTOR

Optional:	
Category	**Abbreviation**
Kommunikationsforen	FORUM
Veranstaltungen/Termine	DATES
Subject Gateways	SUBGATE

Table 1 A proposed list of resources' types and the corresponding abbreviations

III.3. The Location / Library Dimension

For the third dimension of the navigation structure, respectively the **resource's location** / **library that holds it**, there is also no agreed vocabulary in use. The options of this dimension are subject of discussion. The categories of the location dimension could be Berlin, Brandenburg and all. For the library dimension, the options are the names of the partner libraries.

III.4. The Accessing State Dimension

Regarding the fourth dimension of the navigation structure, the **resource's accessing state**, a resource could be **free** – meaning could be accessed by any user, **restricted** – meaning could be accessed just by a user that belongs to an institution which has a license for using it and **free-restricted** – meaning that the resource covers free information and restricted information (e.g. ZDB). At the partner libraries level, there is no agreed vocabulary in use.

IV. The Collection Concept

IV.1. The Motivation

Thinking now about the KOBV users – why would they access the KOBV portal? May be because they are interested in discovering an assembly of data which best fits the topics which they need to investigate (e.g. social science, physics, biology, etc.), or may be the users are interested in discovering an assembly of data which is stored in a certain form (e.g. books, link collections, databases, e-journals, etc.), or may be they are interested in discovering the resources from their surroundings (e.g. Brandenburg, Steglitz, Zehlendorf, etc.).

In overview, the users are interested in discovering subsets of data from various datasets and various locations. These observations drive us to the conclusion that the data need to be organized in a certain way, in an assembled way based on common features (e.g. the covered subjects, the type, the location area, etc.) and the user will begin his "trip" navigating through these assembled data sets. Then the user will "drill down" the assembled datasets in order to find more detailed information. And this provides the motivation for a collection concept.

Another motivation for this concept is based on the number of the resources that should be integrated in the KOBV Informationsportal and the size of these resources. If metadata records from, for example, FU, TU and HU OPACs were browsed, the navigation structure could not be useful due to its size. Instead, metadata about the FU, TU, HU datasets should be browsed. In other words, **the browsing in the KOBV Informationsportal will be performed at the collection level**.

In this context it is necessary to note **the equivalence** between the terms **resource** and **collection**. With this observation, the users of the KOBV Informationsportal, in the discovery process, will navigate through collections administered by different partner libraries. By describing their collections and making available the descriptive metadata, the partner libraries will enable the navigation in the KOBV Informationsportal. At this point, the collection concept needs further explanation.

IV.2. The Definition

According to A.Powell[11], a collection could be defined as “an aggregation of physical and/or electronic items”, with the observations that the items could also be collections, there could be collections of metadata about other collections and an item could be part of more than one collection. Following the same ideas, J.Pete[12] defines a collection as “any aggregation of individual items (objects, resources)”. Also R.A.Bull[13] has defined collections as “a group of related objects and/or other collections which may be centrally located or possibly distributed across locations”.

To note the consistency of the aggregation ideas and of the definition’s recursivity, and the inconsistency of the name of the components of a collection (items or objects). In order to be consistent with the IFLA Functional Requirements for Bibliographic Records[14], the term item will be used in the context of the KOBV information environment. To conclude, a **collection** could be perceived as an aggregation of related items and/or other collections. A collection could be visualized as an inverted tree where leaf nodes are items of the collection and non-leaf nodes are collections.

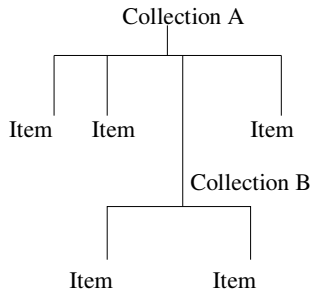

Fig.3 The visualization of a collection

In the above picture, Collection A consists of three Items and a Collection. Collection B, which is a sub-collection of the Collection A, consists of two Items. This should not be

perceived as a requirement for a collection - to have more than one item or to include other collections - a collection could contain just one item. The number of items in a collection is not important and depends on the institution which is responsible for the collection administration.

IV.3. Examples of Collections

As an example of a collection, someone could think about a library catalogue (e.g. http://se1.ub.fu-berlin.de:4505/ALEPH or http://catalog.loc.gov), a link collection (e.g. the links stored by SBB or the link collection stored by the Denmark's Electronic Research Library http://www.deff.dk/?p=main&lang=eng), an Internet directory (e.g. Yahoo), a subject gateway (e.g. SOSIG http://www.sosig.ac.uk or EEVL http://www.eevl.ac.uk), a Web index (e.g. Alta Vista), an A&I database (e.g. PubMed http://www.ncbi.nlm.nih.gov or BIOSIS http://www.biosis.org or INSPEC http://www.iee.org), an e-journal system (e.g. ZDB http://zdb-opac.de), a newspaper (e.g. http://www.tagesspiegel.de), a collection of images, sounds, software, etc.

IV.4. The Granularity of a Collection

Faithfully to the KOBV decentralization policy, the granularity of the aggregation and the relationship between the items that are aggregated in order to form a collection depend on each partner library. Each library can choose between different degrees of aggregation in determining which are their collections. According to Carl Lagoze[15] the aggregations defined by collection items could be: administrative (identifying metadata records aggregated by a distinct entity, such as a library), semantic (identifying a group of items based on factors such as subject relation, age appropriateness) or personal (aggregating items based on a person's idiosyncratic preferences).

There is no exact delimitation between what represents a collection and what represents an item. For example, an e-journal system (such as the Darwin system of the FU Library) could be a collection, in which case the items are the e-journals. But also an e-journal could be modeled as a collection, in which case the items are the articles. Each partner library could decide its collections based on its preferences and needs – the aggregation and granularity of the collections follow a decentralized approach.

But a unified approach should be followed in the case of describing these collections. As it was stated, in order to navigate in the KOBV information environment, the user will navigate through collections. Therefore it is necessary to agree upon a common set of descriptions for the organizational structure that it is layered on top of the data in order to provide the necessary navigational means. Moreover, the collection description could also enable the discovery of the collection for which there is no item-level description.

In summary, for enabling navigation at the collection level in the KOBV Informationsportal, three issues should be addressed: **issue one** - an agreement on a **common set of metadata elements for describing the collections** must be reached; **issue two** - a **collection description service** which enables the partner libraries to describe their collections should be established; and **issue three** - a **collection discovery service** which collects and stores the partner libraries related collection metadata has to be set.

It is worth to be emphasized here that aim of the KOBV-Informationsportal is to be based on a **metadata-sharing process**. Once a resource was described, its metadata should be shared for reusing purposes. From this requirement, a new issue that should be addressed arises: a **collection customization service.** The input of this service should be the metadata generated by the centralized collection description service and the output - an HTML file generated by applying different formatting options to the collection descriptions. The HTML file could then be locally saved and displayed by the partner libraries in the local portals / browsers.

Depending on the number of the collections that will be integrated in the browsing structure, another service could be useful to be implemented – a **collection-level search service**. With the time, if a large number of collections will be described, the browsing structure could become too overwhelming. Therefore a search at the collection-level in the dimensions of the navigational structure could help the user in faster discovering of the collections of interest.

V. Possible Implementations

V.1. The KOBV Collection Metadata Model

As it was stated above, there is a need to agree on a common set of metadata elements for describing collections. The RSLP Collection Description Model[16] and the correspondent Collection Description Schema[17] were chosen as a base for the KOBV Informationsportal collection description metadata element set. Two main reasons have grounded this decision: its generality and its interoperability.

Regarding its generality, it can be pointed out that this model could be applied to collections of all kinds (including library, art and museum materials), e.g. collections of physical items, collections of digital surrogates of physical items and collections of born-digital items. Regarding its interoperability, this research project has been the base for well-known projects in the digital library field such as UKOLN and RENARDUS. Moreover, there is a DCMI Collection Description Working Group with the goal of developing a Collection Description Schema until May 2002 and subsequent deliverables (such as XML Schema, Enumerated list of collection types, Crosswalks/mappings, etc.) until October 2002[18]. In the creation of this Collection Description Schema (which is not yet available), the RSLP Collection Description Schema and the correspondent Analytical Model[19] is cited.

The RSLP Collection Description Schema[17] was adapted to fit the KOBV information environment needs. The resulting metadata elements set is presented below:

KOBV Collection Description Metadata Schema – Proposal		
Name	**Semantic**	**Coding**
Collection related Metadata Elements		
Title	The name of the collection – that will appear in the navigation structure – free text	DC.Title
Identifier	A formal identifier for the collection - URI	DC.Identifier
Collector	The identifier for the institution who gathers the items together – could be standardized (e.g. could be the acronym of that institution FU, TU; HU, etc – an identifier that uniquely identified that institution in the KOBV information environment)	DC.Creator
Short Description	A short description of the collection – could have a standardized structure (e.g. the goal and the subject covered, etc.)	DC.Description
Long Description	A detailed description of the collection – could have a standardized structure (e.g. for restricted collections - what the user should do in order to gain access to it; for offline resources – the address of the collection's location, opening hours, special conditions for access, etc.)	CLD.Note (sub-property of DC.Description)
Subject	The topic of the content of the collection – DDC – could be standardized (e.g. to contain the class number or the class number and the name from the classification schema used by the browsing structure)	DC.Subject
Language	The language of the items in the collection	DC.Language
Physical Characteristics	The physical or digital characteristics of the collection - could have a standardized structure (e.g. online, offline - for CD-ROM, printed – for books, etc.)	DC.Format
Format.Extent	The size of the collection - could have a standardized structure (e.g. about x items)	DC.Format DCq.Extent
Type	The type of the collection – a controlled vocabulary could be used according to Table 1	DC.Type
Access Control	A statement of any access restrictions placed on the collection – free, restricted, or free-restricted	DC.Rights (or CLD.AccessControl)

KOBV Collection Description Metadata Schema – Proposal		
Name	**Semantic**	**Coding**
Date.Issued	Date of formal issuance of the collection	DC.Date DCq.Issued
Sub-collection	The identifier of a second collection contained within the current collection – could be standardized (e.g. the acronym could be used)	DC.Relation DCq.HasPart
Super-collection	The identifier of a second collection that contains the current collection – could be standardized (e.g. the acronym could be used)	DC.Relation Dcq.IsPartOf
Associated collection	The identifier of a second collection that is associated by provenance with the current collection – could be standardized (e.g. the acronym could be used)	CLD.hasAssociation (sub-property of DC.Relation)
Place	The spatial coverage of the items in the collection	DC.Coverage DCq.Spatial
Time	The temporal coverage of the items in the collection	DC.Coverage DCq.Temporal
Postal address	The postal address for the physical location	CLD.Address
Post code (PLZ)	The post code for the physical location of the collection	CLD.Postcode
Collection-Administration related Metadata Elements		
Acronym	The acronym of the collection – should uniquely identified that collection in the KOBV information environment – could be standardized	Ren-cld.Acronym

Table 2 The KOBV Collection Description Metadata Schema – Proposal

The metadata elements from the above metadata schema are needed by the Collection Discovery Service in order to build the navigation structure in the KOBV Informationsportal, but each partner library could add its own metadata elements. The additional metadata elements will not be used by the Collection Discovery Service. For example, for local administrative purposes and/or for the collection description service, the collection description could also include keywords.

It is also worth to emphasize that the proposed KOBV Collection Metadata Schema is mainly Dublin Core based because Dublin Core is a widely spread metadata standard in the digital library community, but other metadata standards could be applied. Because the elements from the DCMES 1.1 and DCMES with Qualifiers namespaces were not enough for encoding all metadata elements, some elements from CLD and Ren-cld namespaces should also be

used. This means that the KOBV application profile for the collections description will be based on four namespaces.

Another possibility would be to define the KOBV Collection Description Metadata namespace comprising all the elements that are not available in the DCMES and DCMES with Qualifiers; for example, in which case the KOBV Collection Description Metadata Schema will be based only on three namespaces. But KOBV should take the responsibility for declaring and maintaining the KOBV Collection Description Metadata namespace[32].

In any implementation-variant, except the administrative elements and the elements Language, Super-collection, Sub-collection, Associated collection, Place, Time, Postal address, all other metadata elements are mandatory. The Subject, Type, Language, Post code, Associated collection, Super-collection and Sub-collection elements could be repeatable. To observe that the accessing state dimension of the navigational structure is not repeatable, whilst the other three dimensions i.e. the postal code (location), the subject area and the type, are repeatable. The decision is based on the cardinality of the relationships between a collection and these dimensions.

It is worth to note that at the KOBV level there exist a simple collection description that is used in the Bibliothekenführer[20]. Therefore, in developing the KOBV collection description metadata schema, in addition to RSLP and RENARDUS metadata schemata, the Bibliothekenführer metadata schema was taken into account. Attention was also paid to the German Virtual Library Application Profile[5], the Kategorien ELSTER[21], and the SSG-FI[22] metadata schemata.

For reaching the interoperability goal, it is also important what encoding schemes are used by the metadata elements. It is worth to emphasize that at this time, there are agreed encoding schemes to be used for some DC elements (e.g. for languages, for format, etc.), but for some elements (e.g. DC.Rights) there is no agreed scheme. Moreover, even in the case of the existing schemes, the values are not appropriate for describing real cases (e.g. DC.Type for the collections from the Berlin-Brandenburg area). But there exists some standards, for example, ISO 8601 (YYYY-MM-DD) for dates which could be used. Also the PND (PersonenNameDatei) schema could be used for individual names (in the order Family Name, Given Name) and the GKD (GemeinsameKörperschaftsDatei) scheme for corporate names – in the case of the collector metadata element.

V.2. The Collection Description Service

The Collection Description Service should enable the partner libraries to describe and administer the holdings that they want to be included in the KOBV Informationsportal. The collection related metadata handled by the collection description service will be called, from now on, **administrative data.**

The service should have a WWW-based interface and should offer **storing and update functionalities**. The storing could be performed in a relational database or as XML files, for example, in the file system.

The service's accessing point could follow a centralized approach i.e. accessing point at KOBV level, or a decentralized approach i.e. accessing point at the partner libraries level. But in any of these approaches, the partner libraries should administer their collections.

In the centralized approach, the Collection Description Service could be an administrative view of the KOBV information portal software. The administrative data will be stored at the KOBV level. In the decentralized approach, the Collection Description Service could be a software tool, preferable based on open source software, provided by KOBV and installed at each partner library. The storing of the administrative data will be performed at partner library level. It is true that the centralized approach does not comply with the KOBV decentralization policy, but it has the main advantages that the administrative data need no Collection Discovery Service and once a collection was described, its description could be further used by other partner libraries that also offer that collection. In the decentralized approach, the administrative data, being stored locally by each partner library, should be first collected at the KOBV level in order to build the browsing structure.

At this time, there are two software tool for collection descriptions – one from the RSLP project[23] and one from the RENARDUS project[24] (which is also RSLP Collection Model based). Both tools use Perl and Apache Web Server and generate XML/RDF descriptions of the collections which are stored in the file system. The RSLP tool has also a version which uses Access relational database for storing, IIS as a Web Server and ASP for dynamic scripts. Both versions of the RSLP tool are available for downloading. In order to rely on public domain software, the first version of the RSLP tool was accustomed to meet the KOBV needs. A prototype of this tool could be reached at http://se3.kobv.de:2727/tool/.

Another possible solution for the Collection Description Service which relies on open software could be the Metadata Tool provided by the Scout Portal[33], which is implemented using PHP, Apache Web Server and MySQL.

V.3. The Collection Discovery Service

The collection discovery service has two main goals: to collect the collection descriptions generated by the decentralized collection description services and to export the collection descriptions stored in a structured form from the partner library into the KOBV-Informationsportal. The difference between these collection descriptions consists of their source: the first ones are generated by the collection description service and the second ones could be generated by another tool (e.g. a content management system). The collection descriptions that are not generated by the collection description service require a normalization process during the export/import process.

The implementation of the collection discovery service is closely related to the Collection Description Service i.e. to the used software tool.

For example, if a version of the above prototype is used, the XML files created by this tool could be harvested from each partner library on a scheduled basis, e.g. weekly. This could

be performed by a Perl script or by using the OAI Protocol for Metadata Harvesting[25]. The harvested data could be stored in a relational database, or as XML files in the file system. A proposal for the export/import service could not be outlined without a closely look on each partner library resources' storing and export capabilities.

V.4. The Collection Customization Service

The collection customization service is necessary only in the case of a centralized approach followed by the Collection Description Service. Its purpose is to enable the libraries to save the centralized collection descriptions locally for displaying. The library should have the possibility to apply some formatting options to a selected group of collection descriptions and save the result as an HTML file. This service could be implemented by retrieving the selected group of collection descriptions as XML documents and applying different XSL stylesheets.

V.5. Pros and Cons of the Navigation at the Collection Level

The main advantage of the Collection Level Navigation approach consists of the ability to provide information about logically defined sets of resources from a broad environment. Furthermore, from an operational perspective, a collection being based just on a logically association, the update of an item metadata, the appearance of a new item or the deletion of an item of the collection does not necessarily imply an update of the collection metadata.

The main disadvantage of the Collection Level Navigation approach could be experienced when the browsing process is transformed into the search process. For example, consider the case of an OPAC that covers Social Science and Language. It is also assumed that DDC is used for the subject area dimension of the navigation structure and the KOBV user is interested in finding resources about Social Science. After discovering the OPAC listed under the subject area Social Science, if the user performs a search in this OPAC, the result set will contain items from the whole OPAC according to the searching criteria, and not only items related to Social Science. In other words, the result set should be limited only to items belonging to that parts of the collections that correspond to the options selected in the navigational structure. But, because of the cardinality of the collection-database relationship and because there is no relationship between the dimensions' values at the item level and at the collection level, the result set could contain items from the whole collections selected to be searched.

Analyzing the collection-database relationship, there are cases when a collection is spread across two or more databases or two or more collections coexist in the same database. Even in these special scenarios, the above metadata schema could be applied. If a collection's items are spread across two databases, the collection could be split in two distinct collections, according to the content of each database, and treated as separate collections. If a relationship is desired to be kept between the two collections, the associated collection metadata element could be used. A more delicate issue appears when a partner library would like to treat a database as more than one collection. In the above example of the OPAC that covers Social Science and Language, having in mind that the subject area is a dimension of the navigation structure, the OPAC (which resides in one database) should be treated as two collections: Social Science and Language. One solution could be to describe the desired collections as two distinct collections, and to use the same Server Name. Or, in this particular case, the collection i.e. the OPAC, could be classify as covering both subject areas (the subject metadata element is repeatable in the KOBV Collection Description Metadata Schema). The advantage of the first solution is that some metadata elements (e.g.

the Short Description, the Long Description, etc.) could contain different explanation for displaying in the navigation structure.

Another possible solution in the case of OPACs that cover all subject areas could be to warn the users that these collections comprise all subject areas. This can be achieved by classifying the collection under the correspondent general class of the used classification schema (e.g. Information & General References in DDC). Another option would be to warn the users about the OPAC's generality in the description that appears in the navigation structure. But all these are possible solutions for integrating the collections in the navigation structure when the cardinality of the Collection-Database relationship is not 1-to-1 and not for better results of the retrieval process.

V.6. Possible Improvements

As it was stated above, the inexistence of a relationship between the dimensions' values at the item level and at the collection level could cause rough results for the retrieval process. This inconvenience could be overcome, broadly speaking, by improving the administrative data with a set of **mapping tables** which should contain the relationship between the collection-level encoding and item-level encoding. From the dimensions of the navigation structure, the most problematic one is the subject area dimension. For this dimension, a better solution could be the use of RDF[26]. The predicate, subject, object triples, that form the basis of the RDF syntax, could be used as a method of representation and storage of thesauri/classification schemas. The RDF data model could express the relationships between terms/classes within a thesauri/classification schema and XML[27] could provide the framework to transport this model.

Another solution could be an **item level normalization process.** The goal of the normalization process is to obtain a common denominator of the item level data. In other words, a KOBV application profile for the item level could be set and each partner library, using a software tool provided by KOBV, should map its data to this profile. During the normalization process, the right values for each item should be incorporated; a right value – meaning the corresponding value for that dimension used in the navigational structure. This approach was followed by the RENARDUS project[29] in developing a broker system. Summarizing, the normalization process of the item-level data involves two phases: firstly, a KOBV application profile for item-level data should be developed, in addition to the KOBV Collection Description Metadata Schema and secondly, each partner library should export its data according to the KOBV application profile. The KOBV application profile could contain the same metadata elements as the KOBV Collection Description Metadata Schema without some administrative metadata elements. In this application profile, it is important that the metadata element Acronym, used by the Collection Description Metadata Schema, to be incorporated in order to know which item to which collection belongs to. A new administrative metadata element is also needed i.e. Full_Record_URL which leads to a detailed display of each item at the originating site. The administrative metadata elements from the KOBV Collection Description Metadata Schema should refer now to these new item stores.

The main advantage of this item-level normalization approach is based on its synergy. Once a KOBV item-level model is agreed on and obtained, a standardization and an enrichment of data is reached. Future projects should not have to deal anymore with the specific heterogeneity of a distributed environment. More than this, the normalization could be achieved in respect to KOBV decentralization policy.

The main disadvantages consist in the allocation of financial and temporal resources required to implement it, on one hand, and the financial and temporal efforts needed to administer the system, on the other hand. There is a first effort required when a partner joins the KOBV and the data should be normalized, and also there is an effort needed further on, caused by the administration of two sets of data. But due to the requirement that the Acronym used for a collection in its description has to be unique in the KOBV environment, the data from one or more databases could be gathered together; and there is no need for each partner to administer duplicate data. It is possible to set Z39.50 Servers that group content data from more partner libraries and/or also to set up an OAI repository (or more OAI repositories).

VI. Conclusions

The KOBV Informationsportal could offer an integrated view of the resources of the partner libraries from the Berlin-Brandenburg area by implementing a four-dimensions collection level navigation structure. For enabling the navigation in respect to the interoperability goal and in compliance with the KOBV decentralization policy, the focus is on agreeing on the options of these dimensions, on setting up a collection metadata schema and on the implementation of the collection services: description, discovery and customization services. While an agreement has been reached regarding the options and the necessary metadata elements, the technical details regarding the implementation of the collection services is still under investigation.

References

1. The Jakarta Project – Jetspeed

<http://jakarta.apache.org/jetspeed/site/>

2. Oracle 9i Application Server Portal

<http://portalstudio.oracle.com/pls/ops/docs/FOLDER/COMMUNITY/OTN_CONTENT/MAINPAGE/KEYFEATURES_BENEFITS/PORTAL_TWP_9_01.PDF>

3. IBM WebSphere Portal Server Architecture

<http://www-3.ibm.com/software/webservers/portal/architect.html>

4. Renardus Application Profile

<http://renardus.sub.uni-goettingen.de/renap/renap.html>

5. **C. Wessel, H. Neuroth**. A Metadata Application Profile for the German Virtual Library, DC-2001, October 2001, Tokyo, Japan

<http://www.nii.ac.jp/dc2001/proceedings/product/paper-10.pdf>

6. Dewey Decimal Classification

<http://www.oclc.org/dewey/>

7. Göttinger Online Classification

<http://www.sub.uni-goettingen.de/ebene_2/2_gok.htm>

8. Hauptklassen der Basisklassifikation – Übersicht

<http://www.gbv.de/du/sacher/bk3_gbv.shtml>

9. KOBV Suchmaschine

<http://search.kobv.de>

10. KOBV Bibliothekenführer

<http://search.kobv.de/bibinfo.html>

11. **A. Powell**. RSLP (Research Support Libraries Programme) Collection Description – Collection Description Schema

<http://www.ukoln.ac.uk/metadata/rslp/schema>

12. **P. Johnston**. NOF-digitise Technical Advisory Service. An Introduction to Collection-Level Description

<http://www.ukoln.ac.uk/nof/support/help/papers/cldintro.htm>

13. **R.A.Bull.** Analysis of the Z39.50 Profile for Access to Digital Collections and the Z39.50 Explain Service for UKOLN

<http://www.ukoln.ac.uk/metadata/cld/study/crossnet/zpadc.pdf>

14. **K.G.Saur**. Functional Requirements for Bibliographic Records,

< http://www.ifla.org/VII/s13/frbr/frbr.pdf>

15. **C. Lagoze, W. Hoehn, and others**. Core Services in the Architecture of the National Digital Library for Science Education (NSDL)

<http://www.acm.org/pubs/>

16. **A. Powell, M. Heaney, M. Lorcan, and D. Lorcan**. RSLP Collection Description, D-Lib Magazine, Sept. 2000

<http://www.dlib.org/dlib/september00/powell/09powell.html >

17. RSLP Collection Description – Collection Description Schema

<http://www.ukoln.ac.uk/metadata/rslp/schema >

18. The DCMI Collection Description Working Group

<http://www.dublincore.org/groups/collections>

19. **M. Heaney**. An Analytical Model of Collections and their Catalogues, Oxford, 14 January 2000

<http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf>

20. **B. Rusch**. Bibliothekenführer

21. **A. Oehler**. Kategorien Elster

22. Categories in the SSG-FI-System

<http://www.geo-guide.de/info/tools/metadata.html>

23. RSLP Collection Description – Tool

<http://www.ukoln.ac.uk/metadata/rslp/tool>

24. RENARDUS Collection Description

<http://renardus.lub.lu.se/cld-tool>

25. The OAI Protocol for Metadata Harvesting

< http://www.openarchives.org/OAI/openarchivesprotocol.htm>

26. Resource Description Model (RDF) model and Syntax Specification

<http://www.w3.org/TR/REC-rdf-syntax/>

27. Extensible Markup Language (XML)

<http://www.w3.org/XML/>

28. DCMI Type Vocabulary

<http://dublincore.org/documents/dcmi-type-vocabulary/>

29. **A. Ardö, C. Doyle, J**. **Hallenberg, and others**. Specification of Renardus Pilot Broker System

<http://nwi.dtv.dk/RENARDUS/D2x/>

30. Regensburger Verbundklassifikation

<http://www.bibliothek.uni-regensburg.de/rvko_neu/>

31. Web Content Accessibility Guidelines 1.0

<http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505/>

32. **R. Heery, M. Patel**. Application Profiles: Mixing and Matching Metadata Schemas

<http://www.ariadne.ac.uk/issue25/app-profiles/>

33. The Scout Portal Toolkit

<http://scout.wisc.edu/research/SPT/demo/>