

## Tilburg University

### Multilingual Access to Subjects (MACS)

Hoppenbrouwers, J.J.A.C.

*Publication date:*  
2001

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Hoppenbrouwers, J. J. A. C. (2001). *Multilingual Access to Subjects (MACS): technical report for the MACS Project*. Unknown Publisher.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Index Data Aps  
Denmark*

*Tilburg University  
The Netherlands*

---

# MACS

## Multilingual Access to Subjects

---

### Architecture of the MACS System

<b>Editor/Author(s)</b>	Jeroen Hoppenbrouwers
<b>Document Number</b>	MACS-2000-01
<b>Date</b>	January 24, 2001
<b>Scheduled</b>	February 2001
<b>Task</b>	N/A
<b>Deliverable</b>	N/A
<b>Status</b>	Final
<b>Version</b>	0.20
<b>Keywords</b>	Link Database, End-User Interface, System Architecture
<b>Abstract</b>	This is a working document. It describes the architecture and design of the MACS multilingual link database management system and the End-User Interface.
<b>Confidentiality</b>	MACS Partners Only
<b>Copyright</b>	© TU/Infolab; January 24, 2001

# Document Change Log

*Parts of the document not yet completed are marked with a • symbol in the text and the right margin. Changes to the previous version are marked with vertical change bars in the left margin.*

**V0.20, 24-Jan-2001** Added the End-User Interface chapter, originally published separately by Index Data.

**V0.10, 30-Aug-2000** Slight notational adaptations in the Zthes record format.

**V0.9, 28-Aug-2000** Modifications in the data schema and a retrospective change of some interface pages, after near completion of the Link Management Prototype.

**V0.8, 27-Apr-2000** New structure of the Zthes record.

**V0.7, 26-Apr-2000** Modifications in the Link User Interface, especially the removal of Delete Link buttons and the tighter definition of the Authorize mechanism.

**V0.6, 19-Apr-2000** Completed the Link Management User Interface, changed some approaches in the management chapter, removed explicit OR relationships, and included a new, revised Zthes record format.

**V0.5, 12-Apr-2000** Added the first drafts of the Link Management User Interface.

**V0.4, 10-Apr-2000** Finished the Link Management User Interface chapter except for the example drawings.

**V0.3, 05-Apr-2000** Started the Link Management User Interface chapter, added a few nasty examples to show the actual work the user interface must do to make the database maintainable. Not finished yet.

**V0.2, 04-Apr-2000** Link Management chapter revision completed. Links are now considered fully bi-directional, ORs are implicit, and ANDs explicit. Replaced *Member* by *Expression*.

**V0.1, 31-Mar-2000** Intermediate release, link model rewritten, but not yet finished.

**V0.0, 29-Feb-2000** Initial release for discussion purposes.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem . . . . .	1
1.2	Proposed Solution . . . . .	1
1.3	Report Structure . . . . .	2
<b>2</b>	<b>Link Management</b>	<b>3</b>
2.1	System Organization . . . . .	3
2.2	Link Semantics . . . . .	4
2.2.1	Single-expression Links . . . . .	5
2.3	Expression Semantics . . . . .	5
2.3.1	OR Semantics . . . . .	6
2.3.2	AND Semantics . . . . .	8
2.4	Subject Heading Semantics . . . . .	8
2.5	Workgroup Organization . . . . .	8
2.5.1	Individual Actions . . . . .	9
2.5.2	Federated Actions . . . . .	9
2.5.3	Conflict Resolution . . . . .	10
<b>3</b>	<b>Link Management Interface</b>	<b>11</b>
3.1	User Model . . . . .	11
3.2	Basic Features . . . . .	12
3.3	Preliminary Sketches . . . . .	13
3.3.1	Search Page . . . . .	13
3.3.2	Search Results Page . . . . .	14
3.3.3	Edit Link Page . . . . .	15
3.3.4	Generic Navigation . . . . .	16
3.3.5	Communication Options . . . . .	16
<b>4</b>	<b>The Link Database</b>	<b>17</b>
4.1	Data Schema . . . . .	17
4.1.1	Design decisions . . . . .	18
4.2	Zthes Record Structure . . . . .	19
4.3	SQL Table Design . . . . .	20
<b>5</b>	<b>User Interface Design and Specifications</b>	<b>23</b>
5.1	Bibliographic Databases . . . . .	23
5.1.1	Bibliographic Database Setup . . . . .	24
5.2	Link Server . . . . .	24

5.3	User Interface Operation . . . . .	25
5.4	Issues and Problems . . . . .	26
5.4.1	Bibliographic Database Capabilities . . . . .	26
5.5	References . . . . .	28

# Chapter 1

## Introduction

### 1.1 Problem

The four partner libraries each maintain an *authority list of subject headings*. These subject heading lists (SHLs) are used to index library databases and therefore must be fairly static to be of practical use. Their different historical development and underlying natural language (English, French, German) make these SHLs unsuitable for cross-databases searches. It is not straightforward to access a library database that has been indexed with subject headings from one SHL while using subject headings from another SHL.

The aim of the project is to provide a way of linking these SHLs so that cross-SHL/cross-database searches become possible, and to develop both a prototype and a management approach to facilitate the maintenance of the links between the “living” SHLs that are, and will stay, in continuous development.

Because the individual SHLs are crucial in the local (in our context, national) database indexing and need to follow the demands of the local librarians and users, it is mandatory that each library maintains *full independent control* over their SHL. There can not be any attempt to centralize SHL (thesaurus) development. The management approach presented in this report acknowledges this given fact.

Lastly, although the current project will deliver only a prototype implementation to show the viability of the management approach, the chosen management solution and database structure must be scalable to international (European) size, embracing tens of thousands of terms and a multitude of inter-SHL links with up to a hundred individual authorities contributing to the link management.

### 1.2 Proposed Solution

The MACS link management system is based upon a federated management approach, where each contributing authority maintains both their own SHL (using the tools they have been using before) and a limited part of the federated link database. This link database is technically centralized, but accessible via the Web. It is therefore possible to manage it in a decentralized fashion. The basic assumption of our approach is:

The links between SHs in the Link Database are provided on a voluntary basis by the authorities, and should under no circumstances be considered

permanent or final.

Therefore, the Link Database can be used by each authority to find links between SHLs, and those links are reasonably up to date and at least somebody found them useful. However, the Link Database is never guaranteed to offer a “hard” mapping between SHLs. This is because of several reasons:

1. All SHLs are in continuous development – they change.
2. Not all SHs have a good mapping to any SH in another SHL.
3. There is no editorial staff that checks and acknowledges the links between SHs which are provided by the individual authorities.

However, there is an analogy to a *peer review process* of the Link Database that makes it possible for links to be improved in time. Chapter 2 expands on this process.

### 1.3 Report Structure

The proposed link management process is described in detail in Chapter 2, including the organization of people and authorities in a federation for mutual benefit. This chapter also describes the negotiation process that takes place between the authorities to establish mutually acceptable links.

Chapter 3 presents the first drafts of a management system that supports the federated approach outlined in Chapter 2.

Chapter 4 contains a technical description, from different points of view, of the actual Link Database and the design of important parts of it.

Chapters 5 and 5 show the architecture and proposed interface of the end user interface that enables the whole system to be treated as one single virtual database, encompassing all individual library databases and subject heading lists.

## Chapter 2

# Link Management

The main purpose of the MACS project is to prove that there is a way of linking up individual SHLs in such a way that:

1. Each SHL maintains its independence.
2. All SHLs can be searched/browsed as a single, big, virtual SHL.
3. The management of the system is scalable and distributed over the participating authorities.

This chapter explains in detail how the management of the system is organized and how the individual SHLs are virtually merged in a federation, with specific procedures for conflict resolution.

### 2.1 System Organization

For the system to work well, it is mandatory that the individual authorities retain full responsibility over their own SHL. This precludes any type of *unified* or *merged database* in which all SHLs are stored together. Such an integrated system would be easier and more consistent to use, because it can be tightly controlled by a central editorial authority, but it is unfeasible for the individual authorities to give up their freedom in changing and managing their own SHL. Previous experiences in database integration have tried to maintain the individual aspects of the contributing databases, but never came to a satisfying solution.

The complete system (see Figure 2.1) will show several distinct databases that are inter-linked. At the periphery of the system are the various SHL databases, which are completely maintained by their respective owners. The MACS system may have read-only access to these databases although this is not strictly required.<sup>1</sup>

At the core of the system is the MACS Link Database, which contains additional data that can be used, on a strictly voluntary basis, to suggest semantic links between various subject headings of the SHL databases. An authority's area of influence encompasses all of its own databases (SHL and the bibliographical databases indexed through the SHL), plus a small part of the centralized Link Database. Importantly, the

---

<sup>1</sup>We still need to decide for the prototype whether or not we will build small databases to emulate the individual SHL databases, or that we will use redundant information in the Link Database for SH names.



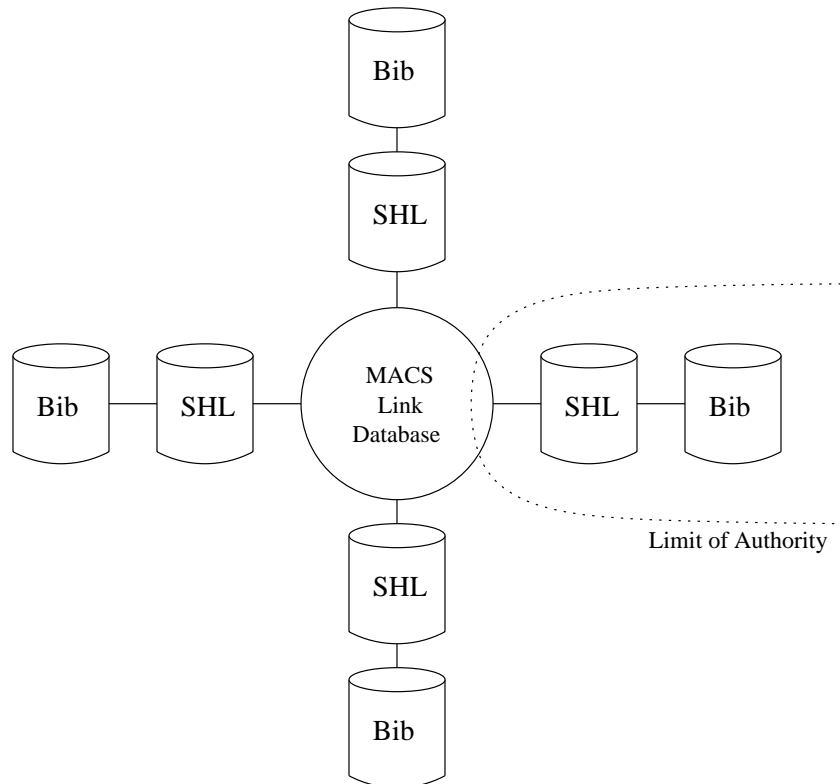


Figure 2.1: Architecture of the MACS System

Link Database *does not contain information that is already available in the individual SHLs*, although for performance reasons some redundancy might be introduced. However, the SHL stays the only authoritative source of information. Browsing the Link Database is no true replacement for browsing the individual SHLs.

The Link Database contains *Links*. A Link is a cluster of *Expressions* (queries, in a sense) in several subject heading languages (Figure 2.2). These Expressions can theoretically be any utterance that would be meaningful in the respective SHL. By placing some Expressions together in a cluster, the system is given the knowledge that these Expressions can be substituted for each other without a *significant* change in meaning of the query. The Expressions, from a user point of view, are interchangeable. Of course each Expression is intended to be used with a specific SHL and underlying database.

Each Expression contains *Terms* (subject headings) from the SHL and operators to connect them, typically AND. The great majority of Expressions will consist of a single Term.

## 2.2 Link Semantics

The prototype will only support one semantical link type, that of *approximate equality*. We restrict the database to this type of link because its only purpose is to provide a 'bridge' between several SHLs. These SHLs may contain more complex semantical

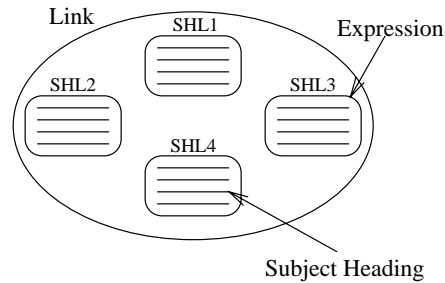


Figure 2.2: A Link between four Expressions

relationships between their subject headings,<sup>2</sup> and it would be a violation of the ‘no redundant data’-rule to add this information to the Link Database. In a sense, the Link Database should contain *no semantics at all*, just blind references to subject heading language expressions in external SHLs.

Scientific applications of thesaurus technology such as (Euro)WordNet<sup>3</sup> also consider the synonymy relationship the most important one, and only add more refined relationships later. Their main semantical building block is the *synset*, a set of semantically (almost) equivalent terms.

In the context of the MACS Project, we say that two SHL expressions are ‘approximately equal’ when:

The expressions contain queries that return the most equivalent results possible, given the information stored in both the databases and the SHLs.

This means that the expressions are not necessarily equal in (linguistic) meaning. We do not consider this a problem. The Link Database is intended to provide a bridge between potentially non-matching SHLs, and under such circumstances any reasonable mapping is better than no mapping at all. Subject heading languages are *not natural languages* and we should emphasize this difference.

### 2.2.1 Single-expression Links

It is a matter of policy whether Links with only one expression are allowed in the Link Database. A Link with one expression does not link anything, so logically it should be disallowed. However, many SHLs contain subjects that do not at all map into other SHLs, and the availability of single-expression links in the Link Database gives a strong clue about the white spots on the map. Filling in the white spots, either by locating appropriate subject headings in other SHLs or by extending those SHLs, is an important management activity, which is much better supported when single-expression links are allowed.

## 2.3 Expression Semantics

Between simple expressions there is no need for expression semantics, since they consist of a single subject heading. However, more complex expressions contain multiple

<sup>2</sup>Semantic relationships such as narrower-term, broader-term, part-of, etc.

<sup>3</sup>textthttp://www.hum.uva.nl/ ewn/

subject headings connected by operators. This section discusses the implications of compound expressions for the organization of the system.

### 2.3.1 OR Semantics

As an example, the following table shows typical one-subject-heading expressions, with each row containing a Link with three Expressions. Note that the equivalence is not *exact*, but only *approximate*. Linguists would likely not accept the equivalence.

LCSH	RAMEAU	SWD
Decathletes	Décathloniens	Zehnkämpfer
Decathlon	Décathlon	Zehnkampf
Discus throwing	Lancer du disque	Diskuswurf
Divers	Plongeurs	Kunstspringer
Diving	Plongeon	Wasserspringen
Hammer throwing	Lancer du marteau	Hammerwurf
Hurdle racing	Course de haies	Hürdenlauf

The previous table contains simple subject headings only, where relationships are straightforward. In the next example, a more complex situation is sketched. There is a significant difference in scope of the used subject headings, leading not to one but to two expression clusters (links) which partially overlap.

LCSH	RAMEAU	SWD
Jumping	Saut en hauteur	Hochsprung
Jumping	Sauts (athlétisme)	Sprung

Here the English (LCSH) subject heading *Jumping* has a broader scope than the French (RAMEAU) and German (SWD) terms *Saut en hauteur/Hochsprung* and *Sauts (athlétisme)/Sprung*. Therefore, when looking for search equivalents of *Jumping*, the system needs to provide two French and two German terms that must be combined by 'OR' in order to cover the whole subject.

While searching with *Hochsprung*, the French term is 'better' (narrower) than the English one, but in the LCSH there simply is no better choice.

It is also possible to *explicitly* combine the RAMEAU and SWD terms into a native SHL expression for the equivalent of *Jumping*. The advantage of this method is that we now have three concepts with clearly separate meanings, instead of deducing a virtual third concept like in the previous example.

LCSH	RAMEAU	SWD
Jumping	Saut en hauteur	Hochsprung
Jumping	Sauts (athlétisme)	Sprung
Jumping	Saut en hauteur OR Sauts (athlétisme)	Hochsprung OR Sprung

Such a link, with more than one subject heading in one of the expressions, is said to be a *compound link* with at least one *compound expression*. Note that leaving the first two lines out and reasoning that the third line contains all information anyway is wrong: the link between *Saut en hauteur* and *Hochsprung* would be lost, and replaced by a more generic link to *Hochsprung OR Sprung*.

### Implicit OR versus explicit OR

We need to make a decision whether to use the implicit or the explicit OR. Advantage of the implicit approach (two links for *Jumping*) is the space saving in the database and the cleanest knowledge representation. Disadvantage is that the system must itself combine the RAMEAU and SWS headings into compound expressions when a user enters the database with *Jumping*.

The explicit approach costs a bit more disk space, and more work to enter. However, it provides the users with more fine-grained tuning capabilities and also makes it clear that there are, actually, *three* concepts involved: the generic jumping, high jumping, and both together. On top, the LCSH also recognizes *broad jumping*. This brings us close to the next step, which would be to also add the knowledge that broad jumping and high jumping are both special cases of jumping. However this would be a bridge too far. We do not want semantical information in the Link Database. In this light, the explicit approach might already be too explicit in terms of semantics.

A possible solution could be to use the expression qualities to prefer some links over others, given the entry point. In this case, the entry points could be sorted by quality.

LCSH	RAMEAU	SWD
Jumping (0.70)	Saut en hauteur (1.00)	Hochsprung (1.00)
Jumping (0.70)	Sauts (athlétisme) (1.00)	Sprung (1.00)
Jumping (1.00)	Saut en hauteur OR Sauts (athlétisme) (0.70)	Hochsprung OR Sprung (0.70)

However, using link qualities for automatic decisions is a fuzzy process and probably should be avoided where possible. Ranking links in a user interface is no problem though. In any case, the system should be able to handle overspecified expressions such as *Hochsprung or Sprung or (Hochsprung or Sprung)*.

There are cases where only one SHL needs an OR operator:

LCSH	RAMEAU	SWD
Runners (Sports)	Coureurs	Läufer
Runners (Sports)	Coureurs	Langstreckenläufer

In such a case, it is more convenient to group the SWD entries together (*Läufer OR Langstreckenläufer*), avoiding duplication of the other SHL entries. But as soon as another SHL is added which also makes a difference between running and long distance running, the link *must* be split into two links like in the previous table. This splitting process (mostly a copy action) will be supported by the user interface. Collapsing a split link into a compound link might also be supported, but is less likely to occur.

The bottom line is that the implicit (split links) approach works, and is probably simpler in the end. The prototype will be built with the assumption that implicit OR will be used. This means that inside expressions, the OR operator will not be allowed, and that link records will need to be partially copied in such cases. But this approach avoids the combinatory explosion of changes when a single new SHL is added that happens to have more specific subject headings available.

### 2.3.2 AND Semantics

For an example of the ‘AND’ operator, look at the following table.<sup>4</sup>

LCSH	RAMEAU	SWD
Child actors	Enfants acteurs	Kind AND Schauspieler
Actors	Acteurs	Schauspieler
Children	Enfants	Kind

It can be seen here that the German SHL cannot express the narrow concept *Child actors* with only one subject heading. Two subject headings must be combined with AND in order to arrive at the intersection of them. This is the reverse operation of the OR-combination in the previous examples. OR is used to combine specialized concepts into a more generalized one, AND is used to make a more specialized concept out of several generic ones.

Unlike the OR-examples, the Link containing *Child actors* cannot be deduced by the system, since the system has no access into the subject headings – it cannot and should not try to parse the meaning of *Child actors*. Therefore, the German expression *Kind AND Schauspieler* must be added manually.

The prototype system will only support one single operator per compound expression, i.e., all subject headings in the compound expression are connected by the same operator. No parentheses, multiple operators, or Boolean ‘NOT’s are allowed, just AND.

## 2.4 Subject Heading Semantics

As said before, the system does not assume *any* semantics for the subject headings. It treats subject headings as completely alien strings, not to be touched at all.

The Link Database will contain unique subject heading identifiers, supplied by the respective authorities, that will refer directly to original SHL entries. There is no technical need to maintain a local copy of the orthography of the subject heading. For system performance and maintenance-friendliness, however, the original subject headings will be copied into the Link Database. The external SHL will forever stay the only true authoritative resource for subject headings.

The Link Database will maintain only one record for each individual subject heading of each SHL, so that changes in the SHL (like a change in orthography) can easily be followed. This centralized record also facilitates the display of all expressions, and therefore links, in which the subject heading participates.

## 2.5 Workgroup Organization

The mandatory independence of the contributing SHLs calls for a federated approach to the maintenance of the Link Database. In such a federation, each partner (authority list maintainer) has the right to pull out at any time, which means that the Link Database must never rely on any SHL being available. It also means that each partner

<sup>4</sup>The subject headings *Actors*, *Children*, *Acteurs*, and *Enfants* have been added to complete the example. They do not necessarily exist in the SHLs, and their addition is fully at the discretion of the authority making the link.

is individually responsible for the maintenance of the links from/to the central (federated) Links and his own SHL. Although collaboration between partners is encouraged, it must be possible to operate the Link Database without such collaboration.

It is completely up to the authorities that maintain a SHL to decide what subject heading ID or IDs and which readable text they put into the Link Database, and in which Link they put it. Of course, if the partners agree on such a standard set of clusters or any form of central (editorial) management, this is possible, but not required. Because the contributing authorities are individually responsible for the links between the federated concept clusters and their own subject headings, they will individually benefit from the work they put into the federated database. The more work they put into it, the better their SHL connects to the others (both ways), so there is an incentive to do it right.

### 2.5.1 Individual Actions

All participating authorities can *add* Links to the Link Database. Authorities are free to create new Links or to add new Expressions to existing Links. It is possible for an authority to add Links between SHLs that are not controlled by him; somebody from the Swiss National Library is allowed to create a link between subject headings of the Bibliothèque Nationale de France and Die Deutsche Bibliothek.

Only controllers of a SHL can *authorize* (or *approve*) expressions containing subject headings of their SHL. Each expression has an independent ‘Approved’ switch that can only be operated by the SHL authority the expression refers to. So, in the above example, the Swiss National Library can add two expressions but cannot authorize any of them. The expression for Die Deutsche Bibliothek can only be authorized by somebody from Die Deutsche Bibliothek, etc. Expressions that are changed by a non-owning authority after authorization are de-authorized and must be reviewed by the owning authority for re-authorization.

Unauthorized expressions are weaker than authorized members. Users can select to see only fully authorized links, where all expressions have been authorized by their respective authorities. This offers them the option to only use the ‘consolidated’ part of the Link Database, and to ignore the unauthorized part that might still be undergoing quality checks.

Authorities are expected to eventually review all unauthorized expressions into or out of their SHL, and where appropriate to authorize them.

SHL controllers also have the right to *delete* any expression in their SHL. If another authority creates an expression that is considered totally unacceptable, it can be deleted by the appropriate authority without further negotiation.

Likewise, it is the authority that has the final word over the subject heading record. Although other parties can create such a record, filling it in with data from the SHL (possibly this is done by the system), only the authority ‘owning’ the subject heading can authorize it. Any change made by others de-authorizes the subject heading record.

### 2.5.2 Federated Actions

The federated model outlined above provides for largely independent updating of the Link Database by the individual partners, where nobody is required to put in work unless they want to gain extra benefits. However, some central management can be added

at the partners' discretion. It is also possible to have more than one Link Database,<sup>5</sup> so that the management can be distributed over several central bodies according to subject field or other criteria. There must be a standardised management interface (maybe up to the user interface level), but there is no need for standardised management between Link Databases.

In case a centralised management effort is called for, organisational negotiations will need to take place about the central body's mandate and task. This body could play the role of an editorial board, with responsibility for the homogeneity of the Link Database. Given the nature of the international collaboration, the cultural differences will likely be considerable, and a wide spread in the levels of ambition and quality assurance by the different partners must be expected. Our approach allows the consortium to experiment with different levels of centralised editorial control, ranging from none to totalitarian. The editorial staff could even double as a telephone/E-mail hotline and consulting service for the different authorities, ensuring that all involved parties share roughly the same ideas of how to do things.

Independent of the availability of centralised management, all federated authorities have access to the full database and can freely add links and delete links 'owned' by them. The central organization offers supervision services in good cooperation with the authorities. The federation model assumes a positive attitude towards cooperation from all members. There is no 'link police' in the current architecture.

### 2.5.3 Conflict Resolution

Despite the open model of cooperation, conflicts between parties will arise. We expect most conflicts to be rather mundane, at the level of (dis)agreeing about the validity of a proposed link. Since each authority can add links to/from other authorities, which means a *proposal* for a link, there will be cases where some authority does not agree with the proposal. The authority to whose subject heading a link member points has the final word over the validity of the proposal, by authorizing or deleting the member.

However, with more than one party involved, possibly including a centralized management body, communication only by means of link addition, authorization, and deletion is certain to fall short. Next to the Link Database but tightly integrated, there must be a discussion and negotiation system through which parties can communicate. It must be easy to add annotations to links and to add comments to these annotations, to escalate discussions about a link into a bigger group, and to moderate these negotiations.

Several standard communication facilities are readily available that can be included in the prototype, such as mailing lists, discussion boards, and live chatting. With the telephone system as a backup, there should be plentiful ways of resolving a conflict without requiring an overruling decision by a commonly accepted authority.

It is an important purpose of the prototype system to evaluate the viability and stability of the federation model, and to suggest possible improvements for the large-scale production system.

---

<sup>5</sup>Several virtual Link Databases can still run on the same server. It is possible to partition a database into smaller virtual databases as well.

## Chapter 3

# Link Management Interface

The main purpose of the Link Management Interface is to offer users (librarians maintaining the SHLs and Link Database) a clear, aggregated, helpful view of the stored links. The relational model underlying the Link Database is far too technical; it is machine-friendly, but definitely not user-friendly. The Link Management Interface is the bridge between the machine-level data storage and the user-level link management. The Zthes communication record is in between, but more on the aggregated link level (user) than on the normalized data level (machine).

The Link Management Interface is not the end-user interface for retrieval purposes. Chapter 5 describes this end-user interface in detail.

### 3.1 User Model

The interface assumes the viewpoint of a user who is responsible for the maintenance of *one particular SHL* and its relationships to other SHLs. Entry into the system is usually with a subject heading from a single SHL, usually the SHL the user ‘owns.’

The system collects all occurrences of the given heading (i.e., every expression in which the heading is an argument) and presents them in a table which shows all Links and Expressions. For a good example, we made up some fictitious subject heading entries to show the various combinations.

LCSH	RAMEAU	SWD
Theater – Financial awards	Théâtre AND Prix et récompenses	Theater AND Kulturpreis
Theater – Non-financial awards	Théâtre AND Prix et récompenses	Theater AND Kulturpreis

The user creating these two links decided to create separate meanings for the LCSH entries, under the assumption that other SHLs would be added in the future that would directly map to both meanings. The prototype supports only this ‘implicit OR’ approach anyway.

When entering the system from the RAMEAU side, using the heading *Théâtre*, the system would produce all RAMEAU headings in which the given heading is the argument of an expression. It is important to note here that *true subheadings are not shown* since the system cannot ‘look into headings,’ and subheadings are a different heading (using ‘–’ within the heading instead of ‘AND’ to connect two separate headings).



Searching for *Théâtre* would retrieve links such as *Théâtre AND Biographies*, but not *Théâtre – Distribution*. The last (sub)heading would categorize alphabetically under *Théâtre*, but not semantically, unless there is a link between them in RAMEAU itself and the Link system is able to tap into this (which it is not). It is a matter of user preference whether (s)he wants to search the Link Database with *Théâtre* (exact matches only) or with *Théâtre\** (alphabetically truncated, to see the subheadings).<sup>1</sup> With just *Théâtre*, the system would produce:

RAMEAU	SWD	LCSH
Théâtre	Theater	Theater
Théâtre AND Bibliographie	Theater AND Bibliographie	Theater – Bibliography
Théâtre AND Biographies	Theater AND Biographie	Theater – Biography
Théâtre AND Prix et récompenses	Theater AND Kulturpreis	Theater – Financial awards OR Theater – Non-financial awards
⋮	⋮	⋮

The list would contain all entries in the Link Database in which *Théâtre* is present in the RAMEAU expression. In the last line of the example, it can be seen that the system automatically collapsed the two links available into one, adding the OR operator between the two LCSH headings. This is a relatively costly operation, requiring argument-by-argument comparison of all expressions of two links. However it will not take more than a second on normal hardware, and since it is not used in the production (high-volume retrieval) system but only in the maintenance user interface and in the Zthes file generator, we expect it will be no problem. The advantage of presenting a collapsed link to the user outweighs the performance loss.

Note that the above list is generated from the RAMEAU point of view. Even if other SHLs provide narrower terms that could be interlinked at a finer level, RAMEAU cannot cope with them, and the list does not show these links. When another SHL is chosen as the view point, and this SHL does contain narrower terms, two or more separate links would show up. The user interface might also have a switch to disable automatic link collapse.

## 3.2 Basic Features

The user interface must at least have the following features to be of any use.

1. User-specific login. No group logins allowed, except possibly for read-only purposes. We need exact authentication and authorization.
2. Selection of SHL for point of view. Default is the SHL ‘owned’ by the person logging in. The SHL used for the point of view is called the *focus SHL*.
3. Standard ‘application view’ with typical search/modify/remove buttons in a concise layout.

<sup>1</sup>It would be possible to add full Boolean (sub)keyword searches if required, but for the prototype, this functionality is assumed to be superfluous.

4. List presentation of found SHL entries in the focus SHL, with the expressions in the linked SHLs next to them, organized per link.
5. Limited browsing options, essentially the list for ‘\*’ or other filters, ordered alphabetically per SHL.
6. Clickable individual expression arguments in all SHLs on display, which take the user directly to the appropriate view of those arguments (focus SHL and search expression implied by the single mouse click).
7. Detail view links that show the administrative details of the various expressions and subject headings.
8. Easy access to modification pages for all relevant items, with individual security features depending on the user’s authorization.
9. Easy access to relevant annotation and negotiation/discussion modules.

Limitations of the HTML Forms interface will likely put constraints on exact layout and functional behaviour (look and feel) of the application. For example, it is very difficult to create self-adapting forms or look-ahead lists with HTML. We try to stick to plain, standard HTML as much as possible, and not use any (semi-)proprietary extensions such as JavaScript or ActiveX. This will ensure the widest browser support available.

In a later stage, 100% pure Java might be used for a two-dimensional browsing interface, but this interface is more useful to display a virtual SHL, created by dynamically combining the participating SHLs.

Also in a later stage, a direct link between SHLs and the database should be established so that there is no need for any re-keying of subject headings. They then can be lifted straight out of the SHL, both as origins and as targets of the link. Such an interface probably requires multiple browsing windows next to each other, which usually precludes the use of plain HTML.

What will *not* be in the prototype is any temporal functionality. There will be time stamps on certain actions, such as the last modification and authorization, but there will not be a complete history of actions. Such functionality is likely required for the production system, but would take up too much development time right now.

### 3.3 Preliminary Sketches

The following sketches are rough outlines of the basic user interface only. Their main purpose is to show the main features and to direct the development of the ‘real’ HTML user interface.

#### 3.3.1 Search Page

Figure 3.1 shows the initial entry interface page. A user (who has already logged in by this time, and therefore is known to the system) selects the required focus SHL. By default this will be the SHL ‘owned’ by the authority the user belongs to. The user can enter either a subject heading or an SHL identifier, the system will search for both. Simple wildcard operators such as ‘\*’ are supported, e.g., for subheading searches. Lastly, the user has to select which SHLs (s)he wants to include in the output listing. This typically will not be the full SHL list, but only a subset, with at least the focus

**macs**  
 multilingual access to subjects

**Search**

[Home](#)

Focus SHL

Search for

Show from
   
 LCSH
   
 RAMEAU
   
 SWD

[Logout](#)

(c) Copyright MACS 2000 For more information, mail macs@cenl.org

Figure 3.1: Search Form

SHL included independent of the user's selection. The default page contains all SHLs in the system.

The 'Search' page is used to directly access the Link Database starting from known terms/identifiers. In many cases, a browsing interface is more likely to offer the required functionality. Browsing can be done with a plain, page-oriented HTML interface, or with a two-dimensional 'Concept Browser' which requires Java support. In all cases, the results are presented on the 'Search Results' page. Note: this is *not* the page that displays the results of a search in a bibliographical database!

### 3.3.2 Search Results Page

After the completion of a search request, the matching expressions in the focus SHL are presented together with their approximate equivalents in the selected SHLs. The focus SHL is positioned in the leftmost column, the other SHLs are ordered alphabetically from left to right. The list can be short or long, depending on the precision of the user's query on the 'Search' page.

All expressions are presented in expanded form, i.e., if for a given focus expression there are several approximate equivalents within any particular SHL, these equivalents will be displayed separately. In Figure 3.2, this has happened with the third and fourth link, which together represent the LCSH subject heading *Theater – Financial awards OR Theater – Non-financial awards*.

By pushing one of the *View Link* buttons, the user can select one particular link to be presented in more detail, at the level of the individual expressions participating in the link. All individual subject headings in the links are also be presented as clickable HTML links, so that users can navigate directly to that term, to another SHL, or to all bibliographical database entries with that term. The exact functionality of these links should be determined in the future; they are not related to the management interface.


		<h1>Search Results</h1>	
<a href="#">Home</a>	<b>RAMEAU</b>	<b>LCSH</b>	<b>SWD</b>
<a href="#">Logout</a>	Theatre Theatre AND Bibliographie Theatre AND Prix et recompenses Theatre AND Prix et recompenses	Theater Theater - Bibliography Theater - Financial awards Theater - Non-financial awards	Theater <a href="#">View Link</a> Theater AND Bibliographie <a href="#">View Link</a> Theater AND Kulturpreis <a href="#">View Link</a> Theater AND Kulturpreis <a href="#">View Link</a>
(c) Copyright MACS 2000		For more information, mail <a href="mailto:macs@cenl.org">macs@cenl.org</a>	

Figure 3.2: Search Results

### 3.3.3 Edit Link Page

The last page in this example (Figure 3.3) show the low-level Link Editor, which is used to modify the expressions that make up a link. Each expression is broken up into arguments, which can be added and deleted one by one. Unauthorized expressions can be authorized by the appropriate people. The *Authorize* button only appears when required, such as in the RAMEAU link.

Each change of any expression in the link will immediately de-authorize all expressions of the link, because the link semantics have changed. All expressions need to be explicitly re-authorized by the appropriate people. A more relaxed model supposes that individual expressions may change over time, but that the meaning of the link at conceptual level stays the same. Changing an expression then would de-authorize that expression only, not all other expressions. The model selection is up to the users.

When the semantics of a link are being redefined, it might be a good idea to temporarily maintain two links, one fully authorized, the other under construction and partially or not authorized. As soon as the new link has been fully authorized and is considered stable, the old link can be deleted. However, the prototype does not explicitly support this operation.

Adding a new term (subject heading) is done by entering the term or its identifier in the text entry field and pushing the *Add* button next to it. The page will be refreshed immediately, now showing the newly added term and a new, emptied text entry field under it. In case a term is not yet known to the system at all, both orthography and identifier must be supplied (the actual system has more form fields than this conceptual drawing).


		Edit Link	
Home	<b>RAMEAU</b>		Created 12-Apr-2000 by J. Hoppenbrouwers
	Theatre AND	<input type="button" value="Delete"/>	
	Prix et recompenses	<input type="button" value="Delete"/>	<input type="button" value="Authorize"/>
	<input type="text"/>	<input type="button" value="Add"/>	
	<hr/>		
	<b>SWD</b>		Created 12-Apr-2000 by J. Hoppenbrouwers
	Theater AND	<input type="button" value="Delete"/>	
	Kulturpreis	<input type="button" value="Delete"/>	Authorized 16-Apr-2000 by T. Place
	<input type="text"/>	<input type="button" value="Add"/>	
	<hr/>		
	<b>LCSH</b>		Created 12-Apr-2000 by J. Hoppenbrouwers
Logout	Theater - Financial awards	<input type="button" value="Delete"/>	
	<input type="text"/>	<input type="button" value="Add"/>	
(c) Copyright MACS 2000		For more information, mail <a href="mailto:macs@cenl.org">macs@cenl.org</a>	

Figure 3.3: Link Editing

### 3.3.4 Generic Navigation

The vertical navigation bar on the left of all pages contains several standard items, such as the depicted *Home* and *Logout* options. These options are available from any place in the system. Other options, not drafted in the example, change according to the page currently on display. They mostly offer ‘Again’ functionality and different views of the same basic information. However, they will never contain functions that change anything in the Link Database. Those functions are uniquely assigned to the buttons.

### 3.3.5 Communication Options

Actions that involve changing expressions of other authorities should be coordinated by interpersonal contact. The system itself does not enforce such communications, but there will be some “intercom” button on most pages to directly access the chat/forum/maillist relevant to the topic.

## Chapter 4

# The Link Database

The Link Database is the technical core of the MACS system. It contains the data required for linking up the several participating subject heading lists, plus some administrative data that is required to manage the system (both technically and organisationally).

Although the Link Database is implemented as a relational database system, the interface to the outside world is not in SQL but in a Z39.50-inspired protocol called Zthes. Most of the user interfacing will be through HTML forms, using a common Web browser.

### 4.1 Data Schema

The data schema of the database, in simplified notation, shows six tables that are interrelated. The base table is the **Links** table, which contains the system identifiers

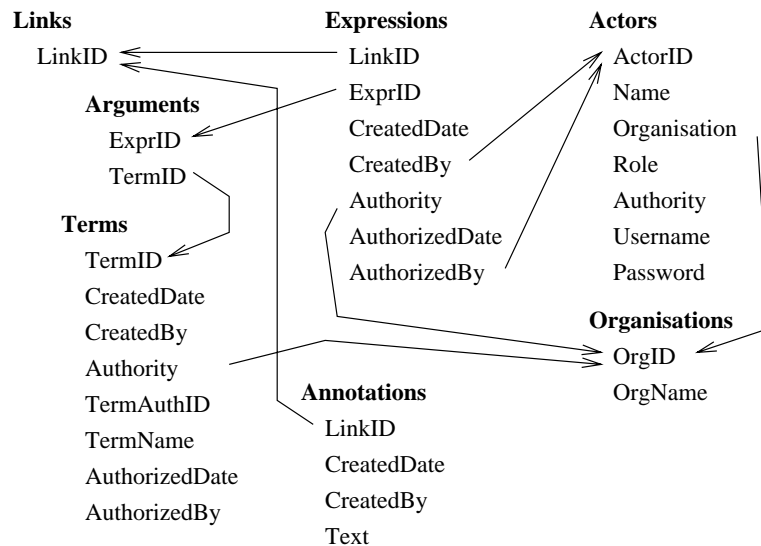


Figure 4.1: Data Schema of the Database

(meaningless for humans) of the links. For the moment we do not really have other elements to store in the Links table, so we could leave it out altogether. In future applications, the **Links** table might also contain data that is tied to a whole link, such as semantical relationships to other links,<sup>1</sup> a canonical link name, etc.

Related to links are **Expressions**, each with their own system identifier as well. A link typically has several expressions. Every expression has several data values, such as the expression creation date and the ID of the person who created the expression. The Authority field contains the identifier of the authority in whose SHL the expression is written. AuthorizedDate and -By fields indicate the date of the authorization of the expression by the authority, and the identifier of the authorizing person. Both fields are empty if an expression is not authorized yet.

Each expression has one or more **Arguments**. The Arguments table only maps from one expression to several terms, it does not add more information. With the help of this table, the **Terms** table can remain completely independent of the Expressions table, we enable easy reuse of existing terms, and we facilitate maintenance.

The actual subject headings are stored in the **Terms** table. This table is by and large a partial copy of the SHL databases. It should not contain any subject heading (term) that is not available in the SHLs. The Terms table also contains typical Link Database information, such as the creation date of the copy of the term (this could also be the date of the last change), the ID of the person originally creating this copy, the owning authority of the term, etc.

All people IDs are referring to the **Actor** table, where all people authorized to use and/or modify the system are represented. It is likely that we need to give the actors roles, e.g., *Administrator*, *Editor*, or *Reader*. Care should be taken that people can leave the operational system, while their name should still be recorded. Therefore the Organisation and Role fields might be empty. Basic items such as username, encrypted password, and ‘owning’ authority are included as well. For the prototype, a decision was made to allow only one authority per actor.

The **Annotations** table simply holds plaintext annotations linked with a Link record. Annotations are part of the discussion mechanism.

Lastly, the **Organisations** table holds all participating organisations. Typically, an organisation is connected to a SHL authority, but this is not a hard link.

Note that this data schema will be invisible to the users. They will get a well-tuned form-like interface to fill in the requested fields, and the inter-table references such as the ActorID will not be visible at all. The users just pull down lists, select choices, and push buttons such as ‘Add’ and ‘Delete,’ and many options such as their own personID will be set at system login time.

### 4.1.1 Design decisions

A few design decisions have been made in the schema that are worth noting.

The decision to give authority IDs to expressions as well as to terms. This is done to rule out the possibility that the same expression contains terms from several authorities (which would be nonsensical). The application should only allow terms of the same authority as an expression to act as arguments.

Likewise, we duplicated the Organisation ID in the Actors table because it is possible that an organisation gives certain update rights to people working for another organisation. Also, when somebody moves from one organisation to another but is still

---

<sup>1</sup>Such semantical links would create a thesaurus out of the Link Database.

allowed to maintain the system, it is handy to have his/her organisational affiliation change. When a person moves and is not longer allowed to access the system, his/her login rights should be taken away, but the person's data should remain in the Actors table, without affiliation to any organisation.

The simplified data schema can and likely will be extended with more fields and possibly tables. We expect that especially the administrative functionality will need to be adapted to the actual way of working of the Link Database maintenance groups. The 'true' Link Database core, however, likely will remain stable and therefore we can fix it in the following section, which describes the record structure in Z39.50-like terminology.

## 4.2 Zthes Record Structure

Whereas the tables contain references to other tables in order to store repeating elements, the Z-like record structure contains the repeating elements in a hierarchical tree format. The record structure is an extension of the Zthes record structure and will be used by the Z39.50 database that contains the link information. A Z39.50 client can search this database for a subject heading from one of the SHLs and the expressions in the other SHLs that are linked to the subject heading are returned as a result of the search, see Chapter 5 for details. A Zthes record relates a subject heading of a given SHL to expressions in the other SHLs. This means that each link in the Link Management System will be represented by one or more Zthes records; each simple, non-compound expression of the link will result in a Zthes record. For notation purposes, we will use a subset of XML to represent the tree format. The system itself is not based on XML, we only use XML as a vehicle to get data across.

The following example in XML representation shows a link between the LCSH-heading "Jumping" and the RAMEAU expression "Saut en hauteur OR Sauts (atletisme)", and the SWD expression "Hochsprung OR Sprung". Most of the administrative information that is used in the link management is left out, since it is not needed in the prototype (see Chapter 5):

```
<Zthes>
  <authority>LCSH</authority>
  <termId>sh85070999</termId>
  <termName>Jumping</termName>
  <link>
    <authority>RAMEAU</authority>
    <exp>
      <operator>or</operator>
      <term>
        <termId>frBN0039853452</termId>
        <termName>Saut en hauteur</termName>
      </term>
      <term>
        <termId>frBN012985577</termId>
        <termName>Sauts (atletisme)</termName>
      </term>
    </exp>
  </link>
  <link>
    <authority>SWD</authority>
    <exp>
      <operator>or</operator>
```



```

<term>
  <termId>041374908</termId>
  <termName>Hochsprung</termName>
</term>
<term>
  <termId>041289900</termId>
  <termName>Sprung</termName>
</term>
</exp>
</link>
</Zthes>

```

The Zthes record is completely self-contained. There is no direct connection between the database structure and the Zthes record structure. Zthes presents the records at a more aggregated level.

The subject heading from the SHL that has the focus is at the top and the structuring follows the Zthes profile. The expressions are included as sub-records in the same way as the terms that are related to the term in the top of the record, are included in relation sub-records (see profile). In this way a Zthes record can express for a given subject heading its relations to other subject headings in the same SHL and its links to expressions in other SHLs. In most cases the internal relations of a SHL and the external links to other SHLs will be in separate databases and thus in separate Zthes records. But in principle the two kinds of information can be merged into one Zthes record.

The expressions themselves are clearly separated from the rest by `<exp>` tags. In order to be able to generate Z39.50 type-1 queries, the element must be structured. The query/expression 'B1 AND (B2 OR B3)' is represented as:

```

<exp>
  <operator>and</operator>
  <term>B1</term>
  <exp>
    <operator>or</operator>
    <term>B2</term>
    <term>B3</term>
  </exp>
</exp>

```

The element `<exp>` is recursive. A DTD can define this recursiveness. The element `<operator>` has two possible values: 'or', and 'and'. In the prototype, 'or' is never stored explicitly in the database, but since the Zthes file is generated relative to a focus SHL, the generator can convert the implicit 'or's into explicit ones.

The simple, non-compound expression that contains only the term S will be represented as

```

<exp>
  <term>S</term>
</exp>

```

### 4.3 SQL Table Design

We make an effort to design the relational system in simple, plain ANSI SQL, so that it is largely exchangeable between database systems. No special features of certain database systems, such as automatic sequences, explicit primary keys, foreign keys, triggers, or constraints are used. This is not the most efficient way to implement a

system, but efficiency is not our concern right now. A true production database would likely be implemented on top of a commercial RDB, and designed by a true database specialist.

```
/* Table structure for the MACS Link Database
   Data types for PostgreSQL */

create table expressions (
  exprID      int4      not null,
  linkID      int4      not null,
  createDate  timestamp not null,
  createdBy   int4      not null,
  authority   int4      not null,
  authorizedDate timestamp,
  authorizedBy int4
);
grant all on expressions to hoppie;
grant all on expressions to www;

create table arguments (
  exprID int4 not null,
  termID int4 not null
);
grant all on arguments to hoppie;
grant all on arguments to www;

create table terms (
  termID      int4      not null,
  createDate  timestamp not null,
  createdBy   int4      not null,
  authority   int4      not null,
  authorizedDate timestamp,
  authorizedBy int4,
  termAuthID  varchar(64) not null,
  termName    varchar(64) not null,
  termSorted  varchar(64)
);
grant all on terms to hoppie;
grant all on terms to www;

create table authorities (
  authID  int4      not null,
  authName varchar(64) not null
);
grant all on authorities to hoppie;
grant all on authorities to www;

create table actors (
  actorID      int4      not null,
  name         varchar(64) not null,
  organisation int4,
  role         varchar(16),
  password     varchar(16),
  username     varchar(16),
  authid       int4,
  lastlogin    timestamp
);
grant all on actors to hoppie;
grant all on actors to www;
```

```
create table organisations (  
  orgID int4 not null,  
  orgName varchar(64) not null  
);  
grant all on organisations to hoppie;  
grant all on organisations to www;  
  
create table annotations (  
  linkID int4 not null,  
  createDate timestamp not null,  
  createdBy int4 not null,  
  text varchar(2048) not null  
);  
grant all on annotations to hoppie;  
grant all on annotations to www;
```

The data types of most ID fields are meaningless integers. Since only the system itself will manipulate IDs, we do not need to take the traditional precaution of using a string data type for numerical codes (e.g., to prevent running into trouble with leading zeros). However, it remains meaningless to do calculations with IDs, with the exception of the (max+1) operation to generate the next valid ID for a specific column.

The `termAuthID` field is a code that can be modified by the user, and therefore the system does not assume any syntax for it.

All date fields got a `timeStamp` type, so that their resolution is not one day, but one microsecond instead. It costs more space but might prove valuable later, when update conflicts etc. might need to be checked.<sup>2</sup>

---

<sup>2</sup>It is usually difficult to implement secure, pessimistic locking in a Web environment, because transactions might span several pages and could be ended without any commitment. The prototype will not pay much attention to these problems.

## Chapter 5

# User Interface Design and Specifications

### 5.1 Bibliographic Databases

All databases are loaded into a stand-alone indexing engine and Z39.50 server, Zebra.<sup>1</sup> The indexing configuration, in general, represent the minimum required to support the prototype application plus basic debugging. Most often, this means that only the

<sup>1</sup><http://www.indexdata.dk/zebra/>

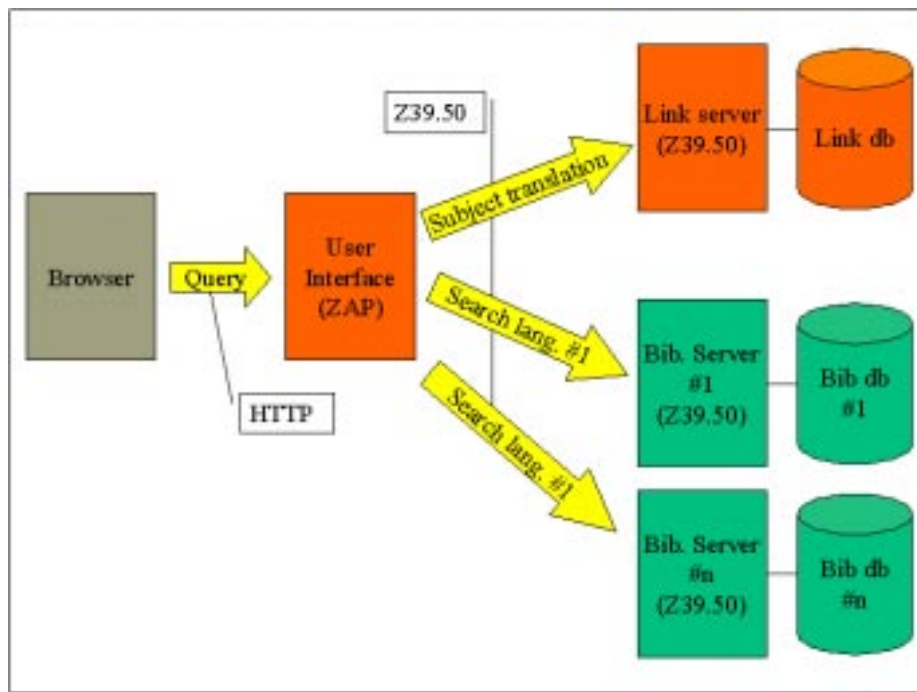


Figure 5.1: Architectural Overview

subject-heading field is indexed (Bib-1 USE attribute 21), although at least the title field is generally available too (Bib-1 USE attribute 4). Searching on the Subject-heading field is carried out either in incomplete-field or complete-field mode, depending on the configuration of the interface. Obviously, “complete-field” mode is most appropriate for the MACS application.

The ISO2709-based records (UNIMARC, USMARC, UKMARC) are generally retrieved for display in their native format and character set, with all fields accessible. The data from DDB is supplied in a text-based format not commonly used in the context of Z39.50. In the prototype, these records are exchanged using the GRS-1 record syntax using local string tags for all elements. The user interface (WWW/Z39.50 gateway) has display filters for each of the involved record syntaxes to realistically simulate a production environment in which each database is hosted by its own, native Z39.50 server.

The Zebra indexing engine makes the data available through a Z39.50 interface. In the prototype, the services Init, Search, and Present are used. In the search, only the complete-subfield, subject-heading attribute combination is used.

In principle, the locally mounted databases should behave roughly the same way as the native databases of the participating organisations might be expected to, and there should be no problem in replacing some, or all of the hosted database samples with access to the full, live system. However, see the section on “Issues and Problems” below.

### 5.1.1 Bibliographic Database Setup

<b>Datab.</b>	<b>Record format</b>	<b>Index setup</b> (subject heading)	<b>Retrieval</b>	<b>Character set</b> (stated by provider)
<b>BL</b>	UKMARC	650/a	ISO2709	?
<b>DDB</b>	?	5100, 5101, 5102	GRS-1	?
<b>BNF</b>	UNIMARC	606/*	ISO2709	ISO 5426
<b>SNL</b>	USMARC	650/*	ISO2709	Z39.47 plus umlauts

## 5.2 Link Server

The link server provides access to the contents of the link database in a framework that is suitable for search and retrieval. The link data made available through the search server is a faithful representation of the data maintained through the link management interface, although for the prototype, there may be a 24-hour delay from when a link is changed to when the change becomes visible through the link server.

The link server is accessed using the Z39.50 protocol, using the Zthes application-support profile. Zthes defines a mechanism for navigating structured thesauri in general. For the purposes of the MACS project, an extension has been defined to handle complex translations of terms.

Through the link server, the link collection is viewed as a set of records adhering to the structure as described in paragraph 4.2 on page 19.



Figure 5.2: Main Window

### 5.3 User Interface Operation

The user interface is realised using a generic WWW/Z39.50 gateway tool called ZAP from Index Data.<sup>2</sup> The gateway is programmed to access two separate types of databases. One type is the Z39.50 front-end to the link database. The second type corresponds to the bibliographic databases.

In the main page of the user interface, the user is asked to enter a query, select a source language, and select one or more databases (Z39.50 servers) in which to search.

When the user enters a phrase, the query is first sent to the Z39.50/Zthes database which hosts the link database. The query consists of two elements one being the term itself, and the other being the source language of the query.

The Zthes database returns a number of matching records, corresponding to matching terms. Matching is done by phrase the search looks for sequences of words in the link database that match the given query.

The user interface extracts the translations that correspond to the SHLs of the selected databases. If the user has selected “browse” on the main screen, he is shown a listing of the matching translations. From this list he can select one or more appropriate translations before proceeding to the actual, multilingual bibliographic search. If the user has selected “Search,” he is sent directly to the bibliographic search, as if he had

<sup>2</sup><http://www.indexdata.dk/zap/>

first selected every matching translation on the “browse” page.

In the bibliographic search, the user interface constructs a query expression for each of the selected databases, and submits the query to that database. The query is a composite of all the selected translations, connected by OR operators (see section on issues and problems). For each term, the query is directed to the subject heading index of the database in question, matching the whole field in each case. The resulting number of hits and the first few records are displayed to the user, and from this point forward the interface behaves like a completely typical Z39.50 client. Records are retrieved in their native format (except DDB, see above), and a MARC format-specific filter provides a user-friendly display independent of the underlying format.

## 5.4 Issues and Problems

### 5.4.1 Bibliographic Database Capabilities

While it is foreseen that MACS will eventually operate in a heterogeneous bibliographic server environment, the MACS prototype itself exists within a controlled environment, where all bibliographic data is mounted on software with well-known capabilities. The nature of the application, however, is such that it may place certain abnormal demands on the servers.

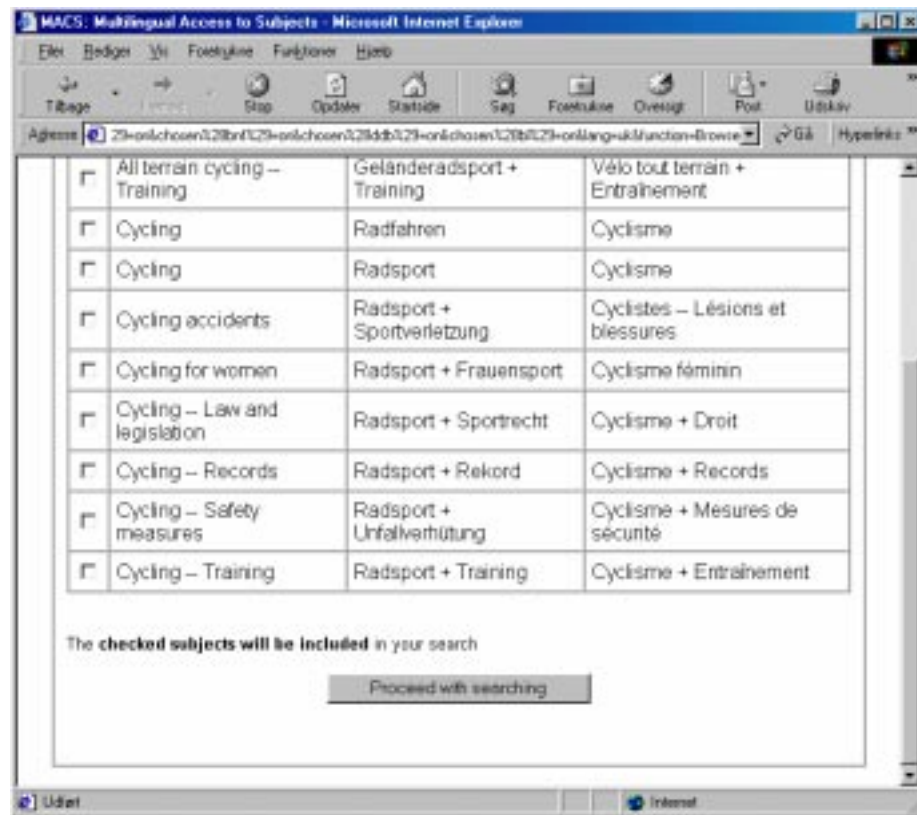


Figure 5.3: Browse Window

For some input terms, the queries generated against a given server may contain a very large number of terms (more than fifty, even in the present, limited prototype), and the resulting query expressions may be several kilobytes long. It is highly likely that some library systems, designed to receive queries directly from users, not from intermediate software, may not be capable of dealing correctly with such complex and voluminous query expressions. There are ways around this, such as constructing the query incrementally executing a series of queries that gradually refine a result set but even such an approach may exceed the capabilities of the bibliographic database, and at any rate is certain to be very time consuming.

This seems to suggest that in a production environment, it will be essential to guide the user towards the correct translation of his query, as opposed to the broad-spectrum set of translations that frequently occur in the present prototype. In a conventional library system, this would correspond to not allowing keyword searches against the subject index but only searches against complete index terms. The present “browse” interface is one approach to guiding the user in this way. A more conventional “index browse” function, which could be made available on the main search page to assist the user in filling in the search field, may be another.

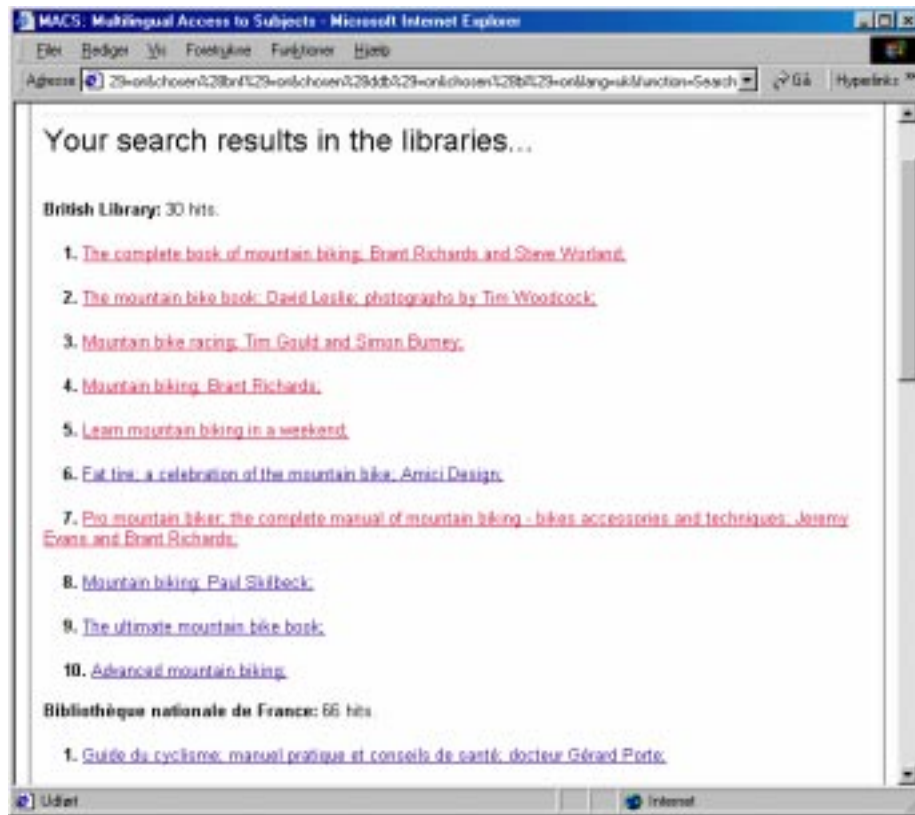


Figure 5.4: Bibliographic Search



## 5.5 References

Z39.50 standard

<http://www.loc.gov/z3950/agency>

Zthes profile for Thesaurus navigation

<http://www.staff.tecc.co.uk/mike/zthes-03.html>

Zebra record indexer

<http://www.indexdata.dk/zebra/>

ZAP WWW/Z39.50 gateway

<http://www.indexdata.dk/zap/>

MACS Prototype direct reference to current version

<http://muffin.indexdata.dk/macs/>

<http://www.loc.gov/z3950/agency>

<http://www.staff.tecc.co.uk/mike/zthes-03.html>

<http://www.indexdata.dk/zebra/>

<http://www.indexdata.dk/zap/>

<http://muffin.indexdata.dk/macs/>