

# Data mining MARC to find: FRBR?

Knut Hegna  
University of Oslo Library  
Knut.Hegna@ub.uio.no

Eeva Murtomaa  
Helsinki University Library  
Eeva.Murtomaa@helsinki.fi

13th March 2002

## Abstract

*In this project MARC data from two national bibliographies is analyzed in the light of the data model presented in the FRBR study from IFLA. The analysis shows that even though the information in the MARC records holds attributes relevant for identifying the work, expression and manifestation entities, the accuracy and formal syntax are too simple to be properly handled by programs. Some of the results may be used to present better hit lists in OPACs. Two suggestions for OPAC user interface based on the ideas of the FRBR study and the results of the project are presented.*

<http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The project and the participants . . . . .	1
1.2	The aim of the study . . . . .	1
<b>2</b>	<b>Mapping MARC fields to FRBR entity identifiers</b>	<b>2</b>
2.1	Identifying the product entities . . . . .	4
2.1.1	Identifying the work . . . . .	5
2.1.2	Identifying the expression . . . . .	7
2.1.3	Identifying the manifestation . . . . .	8
2.2	Works of single personal authorship (AACR2 21.4 A) . . . . .	10
2.2.1	A single work . . . . .	10
2.2.2	Several works . . . . .	11
2.3	Works of shared responsibility (AACR2 21.6) . . . . .	13
2.3.1	One work . . . . .	13
2.3.2	Several works (AACR 21.7) . . . . .	14
2.4	Relationships . . . . .	14
<b>3</b>	<b>Testing the field selections</b>	<b>15</b>
3.1	Henrik Ibsen . . . . .	16
3.2	Jostein Gaarder . . . . .	17
3.3	Dag Solstad . . . . .	18
3.4	Herbjørg Wassmo . . . . .	18
3.5	Tove Jansson (1914-2001) . . . . .	19
3.6	Väinö Linna (1920-1992) . . . . .	20
3.7	Mauri Kunnas . . . . .	22
<b>4</b>	<b>Combining two national bibliographies</b>	<b>23</b>
<b>5</b>	<b>User interface</b>	<b>25</b>
5.1	Presentation fields of the three levels . . . . .	26
5.1.1	The work . . . . .	27
5.1.2	The expression . . . . .	29
5.1.3	The manifestation . . . . .	30
5.2	The card catalogue idea . . . . .	32
5.3	The tree structure idea . . . . .	33
<b>6</b>	<b>Summary</b>	<b>35</b>
<b>7</b>	<b>Testing the methods on BIBSYS data</b>	<b>36</b>
7.1	Herbjørg Wassmo . . . . .	37
7.2	Henrik Ibsen . . . . .	38
7.3	Jostein Gaarder . . . . .	39
7.4	Dag Solstad . . . . .	39

7.5 BIBSYS catalogues: summary . . . . .	40
<b>8 Moving on</b>	<b>40</b>
<b>Appendices</b>	<b>43</b>
<b>A Project description</b>	<b>43</b>
<b>B Example records: single works by Tove Jansson</b>	<b>44</b>
<b>C Example records: several works in a single manifestation</b>	<b>44</b>
C.1 Norway . . . . .	44
C.2 Finland . . . . .	47
<b>D Example records: several works with different authors</b>	<b>50</b>
<b>E Example records: hidden relationships</b>	<b>52</b>
<b>F Example records: Tove Jansson</b>	<b>57</b>
<b>References</b>	<b>62</b>

## List of Tables

1	<i>Attributes identifying the work entity in the MARC formats. . .</i>	6
2	<i>Attributes identifying the expression entity in the MARC formats. . .</i>	7
3	<i>Attributes identifying the manifestation entity in the MARC formats. . . . .</i>	8
4	<i>Results of using the work reduction procedure on records from the Norwegian (n) and the Finnish (f) national bibliographies. . . . .</i>	15
5	<i>The table shows the distribution of expressions and manifestations over the works for the records of the hit list. The ninth row shows that one work had 21 expressions and a total of 28 manifestations, taken from the same number of records. . . . .</i>	17
6	<i>Attributes of high value in order to <b>select</b> a work. . . . .</i>	27
7	<i>Attributes of high value in order to <b>select</b> an expression. . . . .</i>	29
8	<i>Attributes of high value in order to <b>select</b> a manifestation. . . . .</i>	31
9	<i>Results of using the work reduction procedure on records from the BIBSYS catalogue. . . . .</i>	37

## List of Figures

1	<i>The table showing an example of fields and subfields selections used when identifying the work, expression and manifestation entities. Under Work you will find the 100 field (relation to creator) as the first identifying element, next the title fields in sequence of decreasing importance (there are more title field than shown). The field 011 (language code) under expression is a special fix made to harmonize different MARC formats. . . . .</i>	5
2	<i>Card catalogue 1. The card headings are composed of work title and person responsible for the work. If there are many manifestations, we present only the expressions, here represented by document language and manifestation title (see text), original language on the top and then languages alphabetically. To each expression is added the number of manifestations of this expression. . . . .</i>	32
3	<i>The manifestations. The sorting is by publishing year. . . . .</i>	33
4	<i>Card catalogue 2. If there are few manifestations under a work heading, they are presented directly on the card in alphabetical order according to language and within each language, chronologically. . . . .</i>	34
5	<i>Tree structure catalogue. The work titles are sorted alphabetically and are expandable branches. The leaves represent expressions. . . . .</i>	34

# 1 Introduction

## 1.1 The project and the participants

This project started in March 2001 after a short email contact following a question from one of the authors to several email discussion groups in the autumn 2000 asking for implementations of the FRBR model<sup>1</sup> in bibliographic systems.

The project description and agreement are given in appendix A. We will here only mention that Eeva Murtomaa is a librarian at the National bibliography in Finland. She participates in the Standing Committee of the IFLA Section on Cataloguing, and the IFLA UBCIM working group on FRANAR<sup>2</sup>. Knut Hegna graduated in computer science at the University of Oslo in 1980 and has worked for many years with national bibliographic systems. He is now working as a senior academic librarian at the Informatics library at the University of Oslo and as a senior engineer at the National library of Norway. In this project he works on behalf of BIBSYS<sup>3</sup>.

In Norway this project has received two grants from the Ministry of Culture, and a scholarship from the Norwegian Library Association, The Library Bureau and the National Library for Knut Hegna's participation at the IFLA conference in Boston, August 2001.

The project has been presented to The Norwegian Catalogue Committee in two of its meetings. We thank the members of the committee for showing interest in the project.

We would like to thank Mirja Anttila (Helsinki University Library) and Sigrun Ask (University of Bergen Library) who, in the busy last days of the project, accepted to spend time reading through the whole manuscript, providing valuable comments on the presentation and language.

## 1.2 The aim of the study

Many years will pass before cataloguing rules and machine readable formats will support the FRBR model. The questions we asked ourselves were: what can we do in the meantime with the records based on the ISBDs, AACR2 and MARC<sup>4</sup>? To what extent could we extract some of the entities and some of the relations of the FRBR model from the MARC records of our national bibliographies?

---

<sup>1</sup>FRBR - Functional Requirements for Bibliographic Records, see [8]

<sup>2</sup>IFLA - International Federation of Library Associations; UBCIM - Universal Bibliographic Control and International MARC Core Programme; FRANAR - Functional Requirements and Numbering of Authority Records

<sup>3</sup>BIBSYS - a library system serving most university and college libraries in Norway.

<sup>4</sup>ISBD - International Standard Bibliographic Description; AACR2 - Anglo-American cataloguing rules. - 2nd ed., 1988 revision; MARC - MACHine Readable Cataloguing

The reader must keep in mind that we put a question mark at the end of the project title. We did not know what results we would gain. We felt the investigation should be open-ended. At least we expected to identify some problems with the cataloguing rules and the MARC format in this respect.

The main idea in this project is to look at several records with similarities on a certain level (e.g. same author) and analyze the differences. We restrict our study to personal authors as we feel this will also illustrate the nature of the problems with corporate authors and conferences.

The project description states that we were to move from author search to title and subject searches. Unfortunately we have not had time or resources to do this, as we felt that this would mean that the project would not be finished in time. We would prefer to present some results now, rather than all the results at a time not yet known.

We will argue that analyzing hit lists generated by author search alone illustrate fairly well the problems and that the methods we have used may inspire other people (or ourselves) to take the study further.

In this project we wished to investigate OPAC user interfaces which support our findings when it comes to bibliographic structure. In order to do this we would have preferred to add a Z39.50-client<sup>5</sup> in order to test the algorithms with many data sets from the two national bibliographies and data from other bibliographic servers. Again, even though open source program code exists for Z39.50 clients<sup>6</sup>, we have not had the time to include it. The result is that the test data have been gathered from the sources as file downloads and transferred to the project programs as files of MARC formatted records.

Two documents which would have been important to our work are the Interim report on expression-level cataloguing from the JSC<sup>7</sup> format variation working group and Tom Delsey's[4] analysis of the relations between the FRBR model and the MARC21 format. Unfortunately these documents came to our attention too late to influence our work substantially. We recommend these documents to our readers!

## 2 Mapping MARC fields to FRBR entity identifiers

We do not think it is necessary here to describe in detail the FRBR model. What is important to our project is that the FRBR model defines four product entities (*work*, *expression*, *manifestation* and *item*). Each of the entities has a set of *logical attributes* describing it. The entity attributes play different roles when it comes to fulfilling the four user goals set up

---

<sup>5</sup>Z39.50 is a standard protocol describing search and retrieval functionalities for bibliographic databases

<sup>6</sup>See for example <http://www.k-int.com/jzkit/>

<sup>7</sup>JSC = Joint Steering Committee for Revision of AACR.

by the FRBR study (to *find, identify, select* and *obtain*). The entities are connected through a set of relationships which all together complete the comprehension of the bibliographic records and make it possible to navigate in the bibliographic universe.

The FRBR model is to some extent present in the MARC record. A bibliographic record describes both the work and the document (manifestation). Even data describing the expression might be found in the record. The information is sometimes found as strictly defined data both from a semantic and a syntactic point of view. On the other hand some information is found in the fields defined to be non-searchable and is only there as information to be presented when the record is shown to the user. There are few rules governing the data in these fields, mainly in the 500 group. This last category has to be data mined in some simple fashion.

Sometimes the information is there, but may only be understood by the human mind and eye, not by a computer program. Information is entered in order to be searchable either as title or person data, but cannot be identified as e.g. *original* titles or *co-authors*.

As mentioned above, the scope of our analysis is limited to records found when searching for specific persons in the function of author (creator). There are two reasons for this: first, we had a limited amount of resources, this work has mainly been done in the evenings and weekends; second, we felt that this procedure would illustrate well the ideas of the FRBR model.

In an appendix the FRBR study maps ISBD, GARE and GSARE<sup>8</sup> data elements to the logical attributes of the model. In many cases it also maps to UNIMARC fields and subfields. This last mapping was the interesting part for us.

The systems in question use different MARC formats. In Finland the libraries were changing over from FINMARC[9] to a special implementation of MARC21 (MARC21FIN) during our project. We had to stick with FINMARC in this situation. In Norway, the national bibliography uses a subset of NORMARC[10] with some local extensions<sup>9</sup>, while BIBSYS uses their own MARC format.

The Nordic countries agreed in 1973 to follow IFLA's standards for bibliographic description (ISBD) and the Anglo American cataloguing rules (AACR) for recording data in machine readable form. This meant especially the national bibliographies and was a good step forward on the way to the goal of the Universal Bibliographic Control. On the other hand, each country translated and adapted the rules in its own way.

At the outset we would like to convert records from the two national bibliographies into UNIMARC in order to make the processing more consistent

---

<sup>8</sup>GARE - Guidelines for Authority and Reference Entries; GSARE - Guidelines for Subject Authority and Reference Entries.

<sup>9</sup>For short, we will use "NORMARC" in this report

and also usable for others.

It would also be easier to use the concordance table of the FRBR study between the attributes of the entities and the UNIMARC fields and subfields. This appeared to be somewhat difficult as no up-to-date conversion tables were available for the MARC formats in question, and doing this mapping is a project in itself.

In the section 2.1 we will do a more theoretical investigation concerning which fields to use when identifying the entities, and in the sections 2.2 and 2.3 we will look into some problems of using these field selections.

## 2.1 Identifying the product entities

In the next sections we will try to map the various attributes of the entities in the FRBR model to fields and subfields in the MARC formats in question. We do this in order to *identify* the entities. The mapping is restricted to the attributes marked as of *high* or *moderate* value in the FRBR study.

For each entity we present a table with the FRBR attributes in the first column, the mapping to fields and subfields in the two MARC formats in the next columns and finally a column indicating whether we use the attributes for our purpose.

In order to test the identifying qualities of various selections of fields and subfields we have made an analyzing program. This program is described in detail in a separate document[7]. The idea of this analysis is first to perform a search in one or both of the national bibliographies and export the records of the hit lists as file downloads in some MARC format. The next step is to import the MARC records of the hit list(s) into the program.

The analyzing program lets one set up combinations of MARC fields to form identifying strings of the entities on the three levels<sup>10</sup>, see figure 1. First we try to form one or more identifying strings of fields and subfields from each record. The strings are supposed to identify works, and are presented in a sorted list bringing together works with identical strings. We then select one of these identifying strings, and assemble all strings that are identical. From this set of records we form new identifying strings by using another selection of fields and subfields. These strings identify expressions. The same procedure is repeated to form identifying strings for manifestation from a selection of identical expression identifying strings. So a rough description of the algorithm is to do three passes through the data sets, identifying one level of entities in each pass.

When doing this analysis, it is important to have in mind that these data were initially made for another purpose. The data forms the basis for producing printed bibliographies and catalogue cards. In these products we can *see* the result with our own eyes: the works, the expressions, the mani-

---

<sup>10</sup>We do not deal with item entity level in this study.



Work	Expression	Manifestation
100:ah	011:a 041:a	260:c
241:a 500:a 505:a	991:a 700*1:ah	245:a
		300:a
		260:b

Figure 1: The table showing an example of fields and subfields selections used when identifying the work, expression and manifestation entities. Under Work you will find the 100 field (relation to creator) as the first identifying element, next the title fields in sequence of decreasing importance (there are more title field than shown). The field 011 (language code) under expression is a special fix made to harmonize different MARC formats.

festations expressed often in natural language. The structure is expressed by the filing of the cards or by the sorted bibliography with its references. We cannot really expect to find the whole FRBR model by a computer program in these data.

### 2.1.1 Identifying the work

In FRBR table 6.1 (Attributes and relationships of works) attributes that are important for identifying works are listed. Some of these are present in the MARC records from the Norwegian and the Finnish national bibliographies. We combine this information with the information given in FRBR chapter 4.2 (Attributes of a work), the tables in FRBR chapter 7, and appendix A (Attributes of a work). The result is given in our table 1.

To get the work title we look for uniform titles or original titles. We first check if there is a 241 field (original title). The Norwegian national bibliography does not use this field, but the Finnish does. If no 241 field is present, we check the 500 and 505 fields for the text *Original title* in the two languages: *Originaltittel.*, *Orig.tit.:* (abbreviated) or *Originaltitler:* (plural) in Norwegian records<sup>11</sup> and *Alkuteos:* or *Alkuteokset:* (plural) in the Finnish records<sup>12</sup>. The original titles always follow this text. If no original titles are found in the 50X fields on the basis of this test, we move on to other title fields in this sequence: 248 \$h (title proper of a part in Finnish

<sup>11</sup>This holds true in most cases, but during tests we came across the 500 field text *Romanens originaltittel: Heart of darkness* in a record concerning a play based on the novel by Joseph Conrad.

<sup>12</sup>Sometimes these prefixes appear with a lowercase first letter.

FRBR attribute	FRBR value	NORMARC	FinMARC	Selected
title of work	high	500 \$a	241 \$a	yes
		505 \$a	500 \$a	yes
			505 \$a	yes
			248 \$h	yes
		240 \$a	240 \$a	yes
		245 \$a	245 \$a	yes
relation to person responsible	high	100 \$a 70010 \$a	100 \$a \$h 70010 \$a \$h	yes
intended termination	high	?	?	no
form of work	moderate	interpretation of Dewey	008 pos 24-27, 29-30, 33-34	

Table 1: *Attributes identifying the work entity in the MARC formats.*

records) and 240 \$a. The last possible solution is to pick the original title(s) from field 245.

Usually the titles mentioned in the 505 field are repeated in the title added entry fields (745 \$a in Finland, 740 \$t in Norway), but there is no way to decide whether the data in these fields are original titles or not. Some times the fields contain the original titles, some times other title information. Neither 745 nor 740 contain information qualifying the title. In many cases these 74X fields are added entries concerning another work than the work in question, but are related to it in some way.

Sometimes original titles or manifestation titles are given in the content note field (505) with no prefix telling exactly what it is. The text *Contents* (in the respective languages) might be a prefix of title information here, but sometimes the information following is original titles, sometimes not. For this reason we cannot use this information for our purpose.

*Relation to person responsible of the work* we find mainly in field 100, co-authors in 700. There are some problems using the 700 field, these will be discussed below and in the sections 2.3 and 2.3.2.

*Intended termination of work* has no relevance for us, but might be interesting for other kinds of literature than fiction on paper (electronic documents).

*Form of work* would be nice to use even though the FRBR study has given it only moderate value as to identifying works. Barbara Tillett[12] analyzes the cataloguing rules and shows that change of literary form represents a new work. In order to differentiate between a novel and a play *based on* the novel, this attribute is essential. Our study showed that in the Norwegian records one had to interpret the Dewey classification number to get the literary form of the work. Doing this was no problem, but it turned out that

FRBR attribute	FRBR value	NORMARC	FinMARC	Selected
relation to person responsible	high	991 \$a 70011 \$a \$e	700*1 \$a \$h	yes yes
form of expression	high	000 pos 6	000 pos 6	no
date of expression	moderate	?	?	no
title of expression	moderate	?	?	no
language of expression	high	008 pos 35-37	041 \$a	yes

Table 2: *Attributes identifying the expression entity in the MARC formats.*

not all records contained Dewey classification. In the Finnish records the literary form or class to which the work belongs is found in positions 24-27, 29-30, 33-34 of the 008 field. The usage of coded information would be worth studying e.g. for distinguishing different works with identical titles. There were, for instance, two different works with the identical name "Mumin", which were interpreted as one work. In this case the nature of content code might have been helpful to differentiate these works. Here again, not all records contained this information. We decided not to use this attribute.

In this study the *title of the work* and the *relation to person responsible* will be used as attributes identifying works.

Further attributes (e.g. date) of the work we assume to find in the record containing the first published manifestation of the author's home national bibliography (wherever that might be, see section 4).

### 2.1.2 Identifying the expression

The FRBR expression attributes are listed in table 2 with their respective identifying value and their MARC counterparts.

*Language of expression* is of high value in identifying different expressions of a work, but it must be supplemented with *entity responsible for expression*, usually the translator.

Both the Norwegian and the Finnish national bibliography register the translator, but the way they do this varies both over the years, and from format to format.

The Finnish national bibliography does not differentiate between the persons entered in the 700 field, but if the record represents a translation (one may check this by looking at the language codes in 041), the translator is usually found in the first appearing 700 field with the indicators *11*. Unfortunately, sometimes co-author is also entered with *11* (should be *10*), so the

FRBR attribute	FRBR value	NORMARC	FinMARC	Selected
title	high	245 \$a	245 \$a	yes
statement of responsibility	high	245 \$c	245 \$d	
edition	high	250 \$a	250 \$a	
publisher	high	260 \$b	260 \$b	yes
date	high	260 \$c	260 \$c	yes
series statement	high	440 \$a 490 \$a	410 \$a 440 \$a 490 \$a	
form of carrier	high	245 \$h	245 \$z	
extent of carrier	moderate	300 \$a	300 \$a	yes
physical medium	moderate	?	?	
manifestation identifier	high	015? 020 \$a	015? 021 \$a	

Table 3: *Attributes identifying the manifestation entity in the MARC formats.*

identification of translator is not 100% correct.

*Title of expression* is information not covered by today's cataloguing rules, but it is usually the same as the title of the manifestation. We can't rely 100% on this fact (see comments on manifestation title below).

*Date of expression* would be the date of the first manifestation published. This requires a database covering all times and all languages.

### 2.1.3 Identifying the manifestation

According to FRBR the manifestation or item identifier has the highest value for identifying the entity. In the absence of a unique identifier, we need a minimal set of attributes to differentiate one manifestation from another. In most cases the minimum level of attributes comprises title, statement of responsibility, edition/issue designation, publisher/distributor, date of publication/distribution, series statement, and form of carrier. (FRBR p. 85)

*Title.* The manifestation title found in field 245 \$a, is normally the original title or some translated version, but it could also be a collective title, like *Works*, or *Selected works*. We regard this attribute as important in identifying the manifestation.

If we wanted to be more specific, the Finmarc subfields for the title proper by another author (245 \$c), and the number and title of a part (248 \$g, \$h) should be taken into consideration. In the same way the Finmarc subfield (245 \$e) for subsequent statements of responsibility, and the statement of

responsibility relating to a part of a multipart item (248 \$d) should be added (see sections 2.2.1 and 2.2.2.)

*Statement of responsibility.* This information is found in 245 \$c (Norway) and \$d (Finland), and is usually taken from the information on the title page. As there are no rules regarding the syntax of this information, we found it difficult to use as an identifying element and it would not add substantially to the identifying process as author was already in use at the work level.

*Edition.* This element was not used. The combination of year and place of publication makes this element redundant in most cases (in our project). The analysis of editions is problematic, because sometimes different editions are recorded as new records and sometimes edition statements are added to the basic description of a previous manifestation.

Any change concerning the intellectual/artistic content of the work causes a new ISBN number. This new or revised edition is recorded separately into the database.

*Publisher.* This element (260 \$b) appeared to be important, especially when differing between editions published in the same language in different countries (UK/US, Portugal/Brazil, Spain/Mexico).

*Date.* Date (260 \$c) represents an important element when differing between different manifestations of the same expression.

*Series statement.* What holds true for the date, goes for this element also. A work might appear within the same year both as a single work and as part of a series.

*Form of carrier.* The form of carrier is the specific class of material designation, or form of carrier for accompanying material statement in the physical description field. Sometimes this information might be given as a note. It belongs to the group of attributes, which in the absence of a unique identifier or a minimal set of attributes to identify an entity, may serve to differentiate between otherwise identical entities. In this connection, we chose the general material designation (Normarc 245 \$h, Finmarc 245 \$z) to represent the form of carrier. Finnish and Norwegian formats differ from each other in explaining the form of carrier statement. In the table 8 (FRBR: Attributes of high value in order to select a manifestation) we chose the subfield 245 \$h (Normarc), and 300 \$z (Finmarc) as attributes for the form of carrier.

*Extent of carrier.* Field 300 \$a in both databases.

*Physical medium.* This element was not used, and the reason is simple, we overlooked it. In most cases we looked into, change of physical medium also led to a different *extent of carrier*.

*Manifestation identifier.* The manifestation identifier serves to differentiate one manifestation from another. ISBN-numbers are indicated in the field 020 in Norwegian records and in the field 021 in Finnish records. This number might be useful for checking the duplicates automatically. The role of different identification numbers should be investigated in further analysis.

## 2.2 Works of single personal authorship (AACR2 21.4 A)

This category of documents may be divided in two: records containing a single work, and records containing several works.

We included in this category (in these categories) manifestations embodying a work or collection of works for which a single person is responsible. In addition there may be several expressions of the work to which different persons are performing different kinds of activity (e.g. by illustrating, translating, editing etc.)

### 2.2.1 A single work

One would expect this to be the easiest part of the investigation always finding the one and only original title in one of the title fields mentioned in table 1. The problems appeared on certain *practical* solutions, e.g. using the added title entry (740) for whatever purpose needed, usually to make a certain form of the title searchable or as a reference in a printed bibliography (see below).

A work or title authority file or consistent use of uniform title information might have solved this problem.

#### The Norwegian national bibliography

The problem of choosing the right title information may be illustrated by the use of 245 and 740 in reverse roles. *Fru Inger til Østeraad* is the original title of a play by Henrik Ibsen. In modern Norwegian *aa* is usually spelt *å* and the soft *d* is now written *t*. So manifestations of this work nowadays usually appear under the title *Fru Inger til Østeråt*. In the card catalogue this represented no problem, they would all be filed chronologically under the work title.

The problems come to surface when you want to use this information in ways not expected by the cataloguers. Here is the title information from three different records in the Norwegian national bibliography:

**Record 1:**

\*24510 \$a Fru Inger til Østråt  
\*74000 \$a Fru Inger til Østeraad

**Record 2:**

\*24510 \$a Fru Inger til Østeraad  
\*7400 \$a Fru Inger til Østråt

**Record 3:**

\*2451 \$a Fru Inger til Østråt

In the first record the original title is added in the 740 field and not in the 240. This is simply violating the cataloguing rules, adjusting the practice to achieve some effects in the online catalogue.

In the second record an added entry is made of the modern form of the title. This is correct cataloguing.

In the third record no original title at all is given.

How could we possibly decide which is the original/uniform title when the practice is like this? This practice also leaves out the 740 field as a source for finding consistent title information.

### **The Finnish national bibliography**

In the Finnish tradition uniform title (240 \$a ) is mainly used in connection with Raamattu (the Bible) or Kalevala. The original title of a single work is entered in the 245 \$a. With regard to expression, the original work title is mostly entered in the 241 \$a, sometimes in the 500 \$a. If the original title of the work is given as a general note (in the field 500 ), it might also be repeated in the title added entry field 745. Note, that the code of the original language of the work and the classification should relate with each other. If the translation is based on another translation (expression) of the work, which is more familiar to the translator, the information is given in the notes. Sometimes and for many reasons the original title information is missing. Consequently our program will treat expression titles in the field 245 \$a as "work titles".

There are differences in recording multipart works according to Finmarc and other formats. In the Finmarc format the second level title information relating to a multipart work is given in the field 248. The number or designation of the work is given in the 248 \$g and the title of the part in 248 \$h. Usually the title of the part appears also as title added entry information in the field 745. Other information relating to a part of a work may be indicated by repeated fields and by defined indicators.

#### **2.2.2 Several works**

Most of the title fields are not repeatable according to the rules. Several work titles are for this reason repeated within the fields either by repeated subfields or the titles are separated by special delimiting character sequences, usually *space-semi colon-space* as the rules state, but in many instances they are also separated by *dot-space-hyphen-space*. This last sequence cannot be used as a delimiter sequence as it coincides with other uses within the same fields.

### **The Norwegian national bibliography**

When we examine Norwegian records it is apparent that the practice of recording the separate works in a single manifestation has varied over the years. Differences are also due to the cataloguing rules in the way field 245 reflects what actually is on the title page. It will require quite complex algorithms to single them out.

In appendix C we give some examples from the period 1978 to 1993 illustrating the different practices.

From the study of these examples, two main strategies might be used to single out separate titles. The first when the titles are separated by semicolons, and the second when the single titles appear in repeated fields or subfields. This is not 100% correct as manifestation titles which are identical to the work titles will not appear in the list of original titles (see the analysis of the works of Henrik Ibsen on page 16).

### **The Finnish national bibliography**

As mentioned above about the Norwegian records the practice of recording the separate works in a single manifestation varies. How to notice, that we have separate works in a single manifestation? Here seem to be different traditions indeed.

Firstly, there are manifestations (collections) without a collective title. Individual work/expression titles embodied in the collection are repeated in the field for title and statement of responsibility by subfields (245 \$a \$a). When the manifestation comprises two or more individual works by different authors, the work/expression title of the first author's work is given in the field 245 \$a and the titles of the other authors' works in the field 245 \$c.

Secondly there are manifestations (collections) with collective titles containing original works. Usually the collective title appears in the title and responsibility area (245) and the titles of the separate works are included in the content note field (505). Sometimes the title information appearing on the title and statement of responsibility area is to be explained as a collection by human eyes only. This is because the title information may be language dependent (e.g. 245 \$a Dockskåpet och andra berättelser).

With regard to the expressions (e.g. translations), the titles of the original works may be expressed in repeated fields or subfields (241 \$a). In addition the titles of the individual expressions may appear in the content note (field 505). In both fields individual titles might be separated by the delimiting character sequences *space-semi colon-space*, in some exceptional cases e.g. by commas or full stops.

Another way is to add the original titles to the general note field (500), and the titles of expressions (translations) to the content note field (505). The third way is to include both the original and the expression titles in the content note (505).

In addition in many cases the individual titles are given as added entries in the field 745 for giving access to the separate works/expressions embodied in the manifestation. Rarely name title added entries (700 \$a \$h \$t) may be used.

One variation still exists, especially concerning the sound recordings. Individual works embodied in the collection may have identification numbers



of their own. So, the title of the work is given both in connection with the ISBN number (in the field 020) and as added entry information (in the field 745).

Collection of works/expressions as part of a multivolume work: Several individual works/expressions may even be embodied in a part of a multivolume work/expression. In this case besides the name and designation concerning the part, the titles of the separate works/expressions embodied in it are given in the content note, and usually repeated in the added entry fields (as above).

### **2.3 Works of shared responsibility (AACR2 21.6)**

We included in this category works produced by the collaboration of two or more persons chiefly responsible for the creation of the intellectual or artistic content of a work. These works may be result from a collaboration or exchange between two or more authors or from separate contributions of different authors.

#### **2.3.1 One work**

The rules governing the main entry allow only one creator to be entered in the 100 field. Additional creators may be entered in the 700 field. The search systems we have studied join information from both the 100 and 700 fields in the same index, thereby indicating that all persons entered in the 700 field are co-creators of the work. But the 700 field might contain persons who in fact appear in many other functions than author (translator, illustrator, editor and so on). The function might be given in the description (245).

*Function* is optional according to the cataloguing rules concerning main or added entries and the number of functions is very limited. Some use this subfield, some not, supposedly because they want to simplify the cataloguing, but, as we see, with dramatic consequences.

This use of the 700 field leads to a lot of noise when searching for specific authors that play more than one role in their creative life.

#### **The Norwegian national bibliography**

In this bibliography translator and illustrator is usually found in the local field 991 and 990 respectively. In addition, where an added entry is needed in the printed bibliography the information from these fields is duplicated in the 700 field with function in subfield \$e. Other functions (*co-author*, *editor* etc) are entered in the 700 field only.

Names from the 700 field with other functions than illustrator and translator are indexed together with names from the 100 field. This makes it difficult to search for names in the function of author alone and this again makes the analysis more cumbersome, as we shall see.

### **The Finnish national bibliography**

The functions of the authors are given only in the description. Nevertheless there is a kind of logic that helps to differentiate between the principal responsibility, concerning the work and the responsibility concerning the expressions (translator, illustrator etc.) automatically.

If the principal responsibility is attributed to one, two or three persons, their names are given in the subfield \$d of the field 245. The only or the first named person is entered in the main entry heading, field 100.

Added entries for the others are given in the field 700 with the second indicator value 0. This indicator is related to the description concerning the principal responsibility of the work.

The names of the translators, illustrators, editors etc. i.e. persons related often to the expression of the work, are entered in the subfield 245 \$e.

Added entries for these persons are given in the field 700 with the second indicator value 1. Once again, there is a relationship between the description of the expression level (245 \$e) and the added entry information (700. 1).

If the responsibility is shared among more than three persons and principal responsibility is not attributed to any one, two, or three (edited works etc.), the main entry is title (245.1). Once again, the second indicator value 1 of the added entry field 700 indicates that the person mentioned does not have the principal responsibility of the work (even if the name is described in the field 245 \$d).

### **2.3.2 Several works (AACR 21.7)**

We included in this category collections of independent works (or extracts from independent works) by different authors. To this category belong even works consisting of contributions by different authors, produced under editorial direction.

This category represents a real problem in the analysis, obvious from the examples given in appendix D. Varying practice over the years makes it complicated to formulate stringent strategies to single out both the different authors and their works and combine them correctly.

The result is that when analyzing these kinds of manifestations we were only able to use the collective title as the work title.

## **2.4 Relationships**

The attributes are basic tools for the users when they formulate the search query for finding the entity they need. Relationships provide additional information that helps the user in making connections between the entity found and other entities that are related to that entity (FRBR p. 56). Relationships are interesting to display designers and they might give new aspects

Author	Number of records	Number of work lines	Number of unique work lines
Ibsen, Henrik (n)	744	914	220
Wassmo, Herbjørg (n)	149	159	19
Gaarder, Jostein (n)	237	237	14
Solstad, Dag (n)	92	93	35
Kunnas, Mauri (f)	130	133	41
Jansson, Tove (f)	576	595	92
Linna, Väinö (f)	168	168	20

Table 4: *Results of using the work reduction procedure on records from the Norwegian (n) and the Finnish (f) national bibliographies.*

to creating navigation possibilities for the users surfing in the bibliographic universe.

In bibliographic records the relations are mainly indicated in the description. This is a problem, because the textual information is very much language dependent. In the near future this problem may be solved by different kinds of identification numbers, which link different entities with each other. The role of coded information and subject headings and classification is important as well.

What kind of relationships can we find from the descriptions?

Bibliographic records of Tove Jansson's works offer good examples of the variety of "hidden" relationships in bibliographic descriptions and some examples are given in appendix E.

### 3 Testing the field selections

Neither of the systems we were using provided the opportunity to search for author alone. This may seem odd, but it is due to the fact that the search indices included also the 700 field (person added entry).

In the Norwegian national bibliography the role of the person is given in the subfield \$e which is sometimes added when the function is other than author. So the system may limit the search to translator, illustrator and so on, but not to *author*. The problem is that you cannot rely on the added entry being an author if the field lacks subfield e. The reason for adding the entry is often to make it searchable, the cataloguer wants the user to hit this record searching for this specific person's name. The practice then might lead to an addition of the person's name both in the 600 and the 700 fields despite the work is only *about* the person. As we shall see this leads to a lot of noise when it comes to organizing the works of a specific author.

In the Finnish national bibliography there is no explicit function at all in the 700 field, but the function might be deduced from other (descriptive)

fields in the record.

The analysis of the Norwegian records is done by a non-cataloguer, so cataloguers should read this with a critical view. Thanks to Annema Hasund Langballe who answered questions concerning cataloguing details in the Norwegian national bibliography.

### 3.1 Henrik Ibsen

Henrik Ibsen should be well known for his plays. We chose to test the fields selection on this author because we expected to find records covering a large span of time and his works appearing in many different languages.

#### The works

From table 4 we see that there were 744 records where Henrik Ibsen appears as author or as a person added entry. We were able to extract 914 work lines from these records and from these 914 lines we identified 220 unique works. Ibsen has written 26 plays and some collections of poems. Why then, 220 works?

By looking into the records we found that on one side there were records we could accept as representing works: collections of letters from Henrik Ibsen ; compilations of his poems; compilations of his plays. These add up to around 20.

On the other hand there were records which lack information or the information is wrong: collections and selections of plays where the original titles or uniform titles were registered or not identifiable even though they were present (12); 33 records where subtitle was included in the title proper, giving rise to 8 unique works; 3 records with misprints in the original title; 14 records with more or less modern spelling differing from the original title; 10 records with no original titles at all resulting in work headings with foreign language titles.

In addition, when several translated versions of his plays appeared in the same manifestation, the original titles were only given for those plays which had a different manifestation title (that is *Hedda Gabler* and *Peer Gynt* were not repeated as original titles in field 240). For these plays then, some manifestations do not appear under the right work heading but only as “hidden” information under manifestations of other work headings.

One of the main sources of error is the use of the 700 field which is included in the author index. In 87 records *Ibsen, Henrik* was added both in the 600 and 700 field, even though the work obviously was *about* him or his works (in some instances the rules say it should be this way). 8 records were adaptations of his works, 10 works of unknown character where his name was added in the 700 field to make it searchable.

So, when all these records (some of which represent several of these problems) are taken into account, we reach an acceptable number of works.

Number of works	Expressions	Manifestations	Number of records
2	1	1	2
1	2	3	3
1	2	4	4
2	3	5	10
1	6	10	10
1	11	13	13
1	18	24	24
1	20	24	24
1	21	28	28
1	21	31	31
1	24	39	39
1	33	50	50

Table 5: *The table shows the distribution of expressions and manifestations over the works for the records of the hit list. The ninth row shows that one work had 21 expressions and a total of 28 manifestations, taken from the same number of records.*

### The expressions

The main problem of identifying the expressions appeared to be lack of information on original language (19 records). Many records were collocated as being the same expression even though different translators were responsible. This is due to the fact that translator was only entered in the statement of responsibility (245 \$c).

It appeared also that we did not identify other kinds of expressions as records or tapes with his plays even though this should be possible (next version of the algorithm).

### The manifestations

We did not identify any duplicate manifestations appearing under Henrik Ibsen in the Norwegian national bibliography.

## 3.2 Jostein Gaarder

*Jostein Gaarder* started his literary carrier around the introduction in Norway of AACR2. He is an author of fictional works for children and juveniles, and school text books on ethics, religion and philosophy. Some of his works have been translated into many languages. Using the records of his might therefore illustrate the consistency of the database over a shorter time period.

In the Norwegian national bibliography there are 237 records of which *Jostein Gaarder* is responsible for the works in question. When using the reduction algorithm this is reduced to 14 unique works. If cataloguing had been more accurate this number might have been reduced to 13, which is

the correct number! The one record that causes the extra work heading has a misprint in the title (*Bibi* for *Bibbi*).

Differentiating between the expressions in the same language fails in some cases mainly because translator is not registered in an identifiable way or not at all (only Norwegian translators are entered in the national bibliography.).

In the analyzing program we also added the option of generating some simple statistics concerning the distribution of expressions and manifestations over the works. Table 5 shows the distribution for the works of *Gaarder*. The average number of manifestations pr expression is approximately 1,5. We found it interesting, but do not draw any conclusions from this material.

### 3.3 Dag Solstad

*Dag Solstad* is one of most influential contemporary Norwegian authors, whose carrier started in the beginning of the 60-s. His works include novels, short stories and works of literary criticism, but they are not well known outside Scandinavia. When searching for him (as author) in the Norwegian national bibliography we got 92 records in which we were able to identify 93 works. The number of unique works counts down to 35, which is three works too much. The first error is a misprint in the title, the second is due to lack of explicit registration of an identifiable title in a multi volume edition of works of Norwegian authors (see examples in appendix D). The third is a CDROM edition of all his works up to 1995. No single titles are entered in this record, only the title of the CDROM.

There are no special comments concerning the identification of the expressions in the case of *Dag Solstad*, apart from translator not being entered in an identifiable way. His works are only translated into Danish and Swedish and usually they appear only in one edition in these languages. As mentioned, the lack of single titles entered for the CDROM edition, stops identifying this edition as a special manifestation of the works included.

### 3.4 Herbjørg Wassmo

*Herbjørg Wassmo* published her first poems in 1976 and 1977. Later, she has created a number of novels. Her works have been translated into the major European languages.

Searching for *Wassmo*, *Herbjørg* in the Norwegian national bibliography results in 149 records containing 159 of her works. The number of unique works counts down to 19. The correct number according to the records should have been 17. The first erroneous work is caused by the same reason as for Dag Solstad: the title of her work *Dinas bok* not being singled out in a multi volume edition of great Norwegian contemporary authors. Three of her novels have been collocated under one title. Both the single novels and the collective work appear as unique works.

As translators are very seldom explicitly entered in the records the algorithm has not been able to differentiate between editions in the same language translated by different persons.

### **3.5 Tove Jansson (1914-2001)**

Tove Jansson's Moomin characters have conquered the hearts of children and adults in Finland and abroad. According to The Finnish Literature Information Centre 221 works by Tove Jansson have been translated into 26 languages. The Fennica database contains at the moment 720 records of works, collection of works or selections of works with the primary or secondary responsibility for the intellectual or artistic content by Tove Jansson. These works/expressions have been published between 1933 and 2001. The earliest work is written under her pen name Vera Haij. In addition to this the Fennica database contains 45 records of works about Tove Jansson.

Characteristic for the production of Tove Jansson seems to be that the same work/expression may exist as an individual work/expression, and as a component part of a work/expression, or it may be included in the collection of works/expressions. In addition, the number of editions reflects the popularity of Tove Jansson's works. However, the analysis of editions is problematic, because sometimes new records are created for different editions, sometimes the edition statements related to the edition of the work are given in the specific edition statement field (250), and sometimes in the note field for unchanged editions (555). As a result, there is a kind of logic: the editions with essentially the same content as the original work, or the work recorded first in the database, are included in the note field. Because any change of the intellectual/artistic content of the work causes a new ISBN number, revised editions are recorded as new works.

#### **Combining Tove Jansson records from different databases.**

For closer examination a test group of 744 Finnish and Norwegian "Tove Jansson" records was extracted by computer. According to the reduction algorithm there were 158 unique works in 706 records containing 744 separate works. 15 records related to the nonmusical sound recordings.

One reason for this test was to find duplicates from different databases, and to have an idea about the amount of so called "Fennica extranea" material, which does not exist in the Fennica database. Fennica extranea concerns works published outside of Finland but with a Finnish author or translator, and books in Finnish published abroad. Surprisingly indeed, of the 42 records extracted from the Norwegian database 9 matched with records in the Fennica database while 33 existed only in the Norwegian databases. Most of those 33 records described "Mummitrollet"-books which were translated from Swedish, and based on the Japanese TV-series.

Collections with or without a collective title: Human effort is needed

for identifying different kind of collections. In this content the statistics are based on the idea, that the original title and the translated title of the same work are treated as one entity. Original titles are extracted from the 241 fields, and work/expression titles from the field 245.

Collections with a common title: We found 35 work/expression collections (i.e. 61 records) with a common title from the group of 744 records. In addition, there were 8 collections without a collective title.

How much statistics produced by the reduction algorithm are lying? How much manual intervention is needed for getting out the whole "truth"? See examples in appendix F.

### 3.6 Väinö Linna (1920-1992)

In the Finnish national bibliography database we found 168 records with principal responsibility by the Finnish author and academician Väinö Linna (1920-1992). As a result of the reduction algorithm we got 17 works with Väinö Linna as an author of the work, 2 works, in which he exists as co-author and 18 records, where he is the subject of the work.

By looking into the records we realized, that instead of 17 works there are in reality only 11 works. Why this? Due to the lack of the original title information (3 records) two main titles of translations are interpreted as work titles. In addition five individual works exist as parts of a multivolume publication. According to our analyzing programme the collective title of this multivolume monography "Kootut teokset" (= Collected works) was interpreted as a work title.

#### Work: *Tuntematon sotilas* (Unknown soldier)

Väinö Linna made his breakthrough with the novel *Tuntematon sotilas* in 1954. The Fennica database contains 72 bibliographic records of the work *Tuntematon sotilas*, from which 20 records are in Finnish and 52 in other languages.

```
*1001 $a Linna $h Väinö
*2452 $a Tuntematon sotilas $d Väinö Linna
*260 $a Porvoo $b WSOY $c 1954
*300 $a 476, [1] s. $c 8:o
```

Work as a part of a multivolume publication: One of these Finnish language expressions of the work *Tuntematon sotilas* is embodied as part 2 of the multi-volume monograph. According to the reduction algorithm the collective title for all parts of *Kootut teokset* was interpreted as a work title.

```
*1001 $a Linna $h Väinö
*2452 $a Kootut teokset $d Väinö Linna
*248 $g II $h Tuntematon sotilas, 56. p. $y 0002
*260 $a Helsinki $b WSOY
*260 1$a 2000
```



\*300 1\$a 478 s. \$b kuv. \$c 24 cm  
\*745 \$a Tuntematons sotilas

## Expressions

In the FRBR-report translations from one language to another are considered as expressions. If translations are taken into consideration, we noticed, that there are 72 bibliographic records describing 23 expressions of the work *Tuntematons sotilas*.

As a result of the reduction algorithm we found that the English expression of the work "Tuntematons sotilas" was embodied in 5 different manifestations (6 records), the Swedish expression in 6 manifestations (10 records), the Norwegian expression in 2 manifestations (5 records), and the Danish expression in 3 manifestation (4 records). Sometimes the work is translated into some language by different translators. An expression is the specific intellectual or artistic form that a work takes each time it is "realized" Any change in intellectual or artistic content constitutes a change in expression. If a text is revised or modified, the resulting expression is considered as a new expression (FRBR p. 18-19). That's why, if two or more translators have translated the same work into the same language, these translations are considered as new expressions. E.g. there are two expressions in the German language of the work Tuntematons sotilas which appear in 4 manifestations. In the same way the Estonian and Czech expressions are translated by different translators. In addition, these 2 Estonian and Czech expressions appear in 2 manifestations each.

## Manifestations

According to FRBR the boundaries between one manifestation and another are drawn on the basis of both intellectual content and physical form. In addition, where the production process involves a publisher, producer, distributor, etc., and there is a change in publisher, etc., the resulting product may be considered a new manifestation (FRBR p. 21).

We used e.g. publisher information (260 \$b) to differentiate the manifestations from each other. The Finnish expressions of the work Tuntematons sotilas are mainly published by the same publisher: WSOY. Some manifestations are distributed by the Suuri suomalainen kirjakerho.

Some manifestations could be separated with the series statement, but we did not use this field.

\*1001 \$a Linna \$h Väinö  
\*241 \$a Tuntematons sotilas  
\*2452 \$a Ukjent soldat \$d Väinö Linna \$e overs. av Odd Bang-Hansen  
\*260 \$a [Oslo] \$b Norske bokklubben \$c 1964  
\*300 \$a 383, [1] s.  
\*440 \$a Bokklubbens böcker \$v 38 \$y 0038

```
*1001 $a Linna $h Väinö
*2412 $a Tunteaton sotilas
*2452 $a Ukjent soldat $d Väinö Linna $e overs. av Odd Bang-Hansen
*250 $a 3. oppl
*260 $a [Oslo] $b Den norske bokklubben $c 1968
*300 $a 383, [1] s.
```

In addition we could not use the language based information from the description, which would be useful for the users.

```
*1001 $a Linna $h Väinö
*2412 $a Tunteaton sotilas $e kuvituksena
      17 taiteilijan TK-kuvia jatkosodan ajalta
```

### **Täällä Pohjantähden alla (Under the North star)**

This work (consisting of three individual parts) is realized through 11 languages and published as one volume (40 records) or separate parts. These parts may be published separately with their own titles (30 records) or as a part of the collection with common title *Kootut teokset* - Collected works - (3 records).

Once again, we separated different manifestations according to the publisher information. E.g. Finnish, Swedish, Estonian, and Russian records appear in two manifestations. In Norwegian records there were 2 translators, but only one publisher. If we consider this trilogy "Täällä Pohjantähden alla" as one work, the expression in Norwegian language appears in one manifestation (even if existing in three separate parts).

### **3.7 Mauri Kunnas**

The works of Mauri Kunnas have been translated into many languages. Using the records of his works might therefore illustrate the consistency of the database over a shorter time period. His works are published between 1978 and 2001.

There are 133 records in which the principal responsibility is attributed to Mauri Kunnas. In 93 records Tarja Kunnas is presented as a co-author. We found Mauri Kunnas as the subject of the work only once.

Via reduction algorithm we got 39 unique works (in reality only 36 works), expressed in 15 languages. Almost all works are written in Finnish. There is one bibliographic record only in Danish, written by Robert Fisker and illustrated by Mauri Kunnas. Four Ricky, Rocky, and Ringo-books are originally written in English and translated into 5 different languages.

Why in reality 36 works? The titles of collections with collective titles were interpreted by machine as new works. In reality the collections *Mauri Kunnaksen koiramäki* and *Mauri Kunnaksen yötärinoita* contained works, which were published separately. The different Ricky, Rocky, and Ringo

books are also published as a collection with the collective name *Rikun, Roopen ja Ringon seikkailut*.

As an example of "hidden works" the material from the cartoon album and from the periodicals published for young people in "Intro", "Help", and "Suosikki" in 1975-86 should be mentioned. This material is included in the collections "Nyrok city kokoelma", and "The best & the worst of Nyrok city : giant gollexon by Mauri Kunnas".

The amount of translations reflects the popularity of the works by Mauri Kunnas. Besides works/expressions in Finnish we found 17 expressions in Swedish, 16 in English, 9 in Italian and Japanese, 8 in German and Icelandic, 5 in Norwegian, 2 in French, Estonian, Latvian, and Danish, 1 in Dutch, Welsh, and Hungarian.

There are 20 records of the work *Joulupukki* (Santa Claus) in the Fennica database. The work is available in 12 languages. When taking different translations and translators as well as different publishers into account, we found that there are 17 expressions of the work *Joulupukki* appearing in 21 manifestations. There are different variations of the English title of the work *Joulupukki* like "Santa Claus and his elves", "Santa Claus : a book about the doings of Santa Claus and his brownies at Mount Korvatunturi", and "Santa Claus : a book about Santa Claus and his assistants at Mount Korvatunturi, Finland". This means, that 3 English expressions of the work *Joulupukki* appear as five manifestations. In the same way three French expressions of the work *Joulupukki* appear in four manifestations, and one German expression appears in two manifestations.

48 Mauri Kunnas records contained edition information. This is not taken into consideration in our study, but the edition information tells something about the popularity of the different works. The winner is again *Joulupukki*. There are 13 Finnish editions between the years 1981 and 1998, and 40 editions of the *Joulupukki* in Japanese between the years 1982 and 1999. Besides this some publishers have several editions of the English translations of the *Joulupukki* (e.g. there are 10 editions between the years 1981-2000 published by Otava, 2 editions published by Methuen Children's Books in 1982, 2 editions published by Harmony Books in 1982 and 1983, and one edition by Crown Publishers in 1985).

## 4 Combining two national bibliographies

### The origin of the record

Unfortunately we did not have time to test the use of the "origin of the MARC record"-information in the Z39.50 environment. As the cataloguing practice and MARC formats differ, the question of the source of the record is important in the global environment, both with regard to the exchange of cataloguing records and with regard to the possibility to do the kind of

analysis we are doing in this project. Interchanging information according to UNIMARC might be a solution, but not all systems provide this.

Actually, there are several control numbers and identifiers in the Marc formats for serving this problem. For example the field 016 “National Bibliographic Agency Control Number” in the MARC21 format. The field 001 contains the control number assigned by the organization creating, using or distributing the record. In addition, the Marc code for the agency whose system control number is present in the 001 field is given in field 003 (Control Number Identifier). For example, the control number identifier Fi-Fenn in the field 003 tells another library that the record’s origin is the Finnish National Bibliography.

An organization receiving a record may move the incoming control number from field 001 (and the control number identifier from field 003) to field 035 (System control Number), and place its own system control number in field 001 (and its control number identifier in field 003). In addition, the code<sup>13</sup> or name of the cataloguing agency that created the original record could be added to the field 040 (e.g. in MARC21Fin).

### **One bibliography supplementing the other**

One idea in this project was to search for a Finnish author in the Norwegian bibliography (or vice versa), identifying the works and Norwegian expressions and then move on to the Finnish national bibliography to fetch other expressions and possibly also further attributes of the work itself by looking at the first published manifestation there.

We looked for ways to identify one author’s home national bibliography, but we did not find any. The language code could not be used as many authors write in other languages than the main language of their own country, many authors publish documents in other countries than their native one.

The Norwegian national bibliography in some instances adds a country code to the author’s name (e.g. *fi* for Finnish authors), but as this practice is far from completely carried out, we could not rely on it.

Maybe, in the future, we could get this kind of information from an authority file, but then, again, people are moving around from country to country and do not *belong* to a specific national bibliography. The work, though, the first published manifestation, belongs to a certain bibliography, so maybe the problem might be solved by the use of a work authority file.

### **Language dependent texts**

Another problem in the kind of analysis we are performing, is language dependent texts used to qualify information in the records. These texts stem from the description and note fields in particular. In the search for original

---

<sup>13</sup>MARC Code List for Organizations, maintained by the Library of Congress contain codes for several organisations.

titles in the note fields we had to look for different texts (*Originaltittel* and *Alkuteos*) to get hold of the right information. This way of qualifying information should be replaced or supplemented by formalized and language independent procedures in the future.

## 5 User interface

The project agreement states that we were to base our suggestions on the IFLA document *Guidelines for OPAC displays*. This document has been withdrawn and as the new version has not yet been published, we felt free to come up with our own proposals.

We will in this section present two different user interfaces based on the structure we find in the records of the hit lists. As the scope of our analysis is limited to search for specific authors, the user interfaces are directed to present their works and the various editions of the works in a structured way.

We believe that the hit list in a search should appear according to the search performed and the results themselves. When you search for a distinct person, the hit list should either consist of his or her works in some order, alphabetical perhaps or chronological or by a list of the different functions he plays related to the entities (author, illustrator, translator).

What we would like to do is to present something like the structure of the card catalogue as *James Duff Brown* proposed in his book *Library classification and cataloguing*[2] in 1916. His arrangement for authors who have written extensively goes like this:

1. **Single works in chronological order of publication**
  - Originals, including manuscripts
  - Reprints
  - Translations in foreign languages
  - Parodies
  - Criticism, etc., of single works
2. **Collected works in chronological order of publication**
  - Author's editions
  - Editor's editions
  - Selections
  - Paraphrases and condensed versions
  - Dramatic versions of works
  - Musical settings of works
  - ...
  - Bibliography
  - Index

Clearly, we lost something moving from the card catalogue and the printed bibliography to the online catalogue! The insight and logic of a human being.

Nowadays, one usually does not see this kind of structure in the hit lists of the OPACs. In some catalogues one might see a structure based on uniform

titles, e.g. a search for the author heading *Ibsen, Henrik* might lead to a hit list resembling this:

```
Ibsen, Henrik [1841-1906]
...
Dukkehjem.
Dukkehjem. English. Doll's house
Dukkehjem. German. Nora
Dukkehjem. German. Puppenheim
...
```

In Norway and Finland there is no tradition for using uniform titles this way in the online catalogues.

We have in the online catalogue no other choice than arranging the works by sorting the original titles, even if it is a collection or selection. We have the opportunity to collocate all works with the same title and present it as one heading. We may at further collocate the different expressions under the same work heading, whether it is a single work or a uniform standard heading like *Collected works* or *Selected works*.

Having done this, one might collocate equal expression identifiers under one unique expression header.

In her book Svenonius[11] writes:

*The essential and defining objective of a system for organizing information, then, is to bring essentially like information together and to differentiate what is not exactly alike. (p.11)*

This is true, but not the whole truth. The objective of differentiating must be understood on two levels.

First, to differentiate between information which is a part of the collocating process. This calls for normalization of data through the use of authority files, consistent registration and so on.

Second, having collocated records on this basis, the system should give enough information to make it possible to differentiate between the records seen from this level. This is about detailing the information and deciding what kind of information in fact differentiate the collocated records *to the benefit of the user*. This is the hard part, because who knows what the user needs? The cataloguing rules made to meet the objectives of the library catalogue tell us what is most important.

## 5.1 Presentation fields of the three levels

When deciding which fields of the entities to present to the user we found it relevant to look at the attributes and relations having high value in order to *select* between the entities in question. These attributes and relations are listed in tables 6.1 to 6.3 in the FRBR study under the column heading *Select*.

FRBR attribute	NORMARC	FinMARC	Selected
title of work		241 \$a	yes
	500 \$a	500 \$a	yes
	505 \$a	505 \$a	yes
		248 \$h	yes
	240 \$a	240 \$a	yes
	245 \$a	245 \$a	yes
form of work	inter-pretation of Dewey	008 pos 24-27, 29-30, 33-34	
relation to person responsible	100 \$a	100 \$a \$h	yes
	70010 \$a	70010 \$a \$h	
referential successor			
referential supplement			
referential complement			
subject entities	6XX, 082 DDC	6XX, 080 UDC	no

Table 6: *Attributes of high value in order to **select** a work.*

To make selection easier it is important also to decide what kind of ordering is most appropriate. The sorting (filing) principle should be easily identifiable by the user and should vary according to which entity level is presented.

### 5.1.1 The work

The presentation of works must depend on how the search was formulated. If you have searched for one specific person, it would be sufficient if the first level of the hit list consisted of the work titles alone, as these would be differential. On the other hand, if the person in question plays a lot of functions to the works, expressions and manifestations, the list should make this clear and let the user decide whether he will proceed with this person as author, illustrator, translator, composer, performer, or whatever or maybe *all* functions<sup>14</sup>.

In table 6 we have listed the FRBR attributes of high value for selecting a work and the relevant MARC fields in the two formats. All these have to be taken into considering in a complete system. If our scope is searching for one specific author, the list might consist of work titles alone, but if the search is for titles the *person responsible for the work* must also be added,

<sup>14</sup>Some Norwegian systems work this way, but neither the National bibliography nor the BIBSYS system do.

as illustrated in figure 4 and figure 5. We have chosen to include both these elements in our experiment, even though our tests only deal with specific authors.

*The form of work* is the class to which the work belongs e.g., novel, play, poem, essay, biography, etc. The form of work may exist also as an addition to uniform title or as a note on nature, scope, literary form etc. We did not use the information codes in our programs because they were not reliable.

The next three attributes in the tables 6 and 7 concern *referential successors*, *referential supplements* and *referential complements* of works and expressions. They reflect relationships between one work and another. This information is usually included in the description, which is very much language dependent, and problematic for machine processing. It is important to note, that referential successors, supplements and complements of the works/expressions are highly dependent on the content of the other works. (FRBR p. 68)

*Referential successor.* According to FRBR a sequel is defined as a referential successor of the work or expression. It is complete in itself, but continues an earlier work, and is very closely connected to it. A sequel has little value outside the context of that other work. The successor type of relationship involves a kind of linear progression of content from one work to the other (FRBR p. 67).

*Referential supplement.* According to the FRBR referential supplements of works and expressions are usually issued separately. They complement already published originals by bringing up-to date material or otherwise continuing or supplementing the original. "The supplement relationship type involves works that are intended to be used in conjunction with another work. Some of these, such as *indices, concordances, teachers' guides, glosses, and instruction manuals for electronic resources* will be so closely associated with the content of the related work that they are useless without the other work. Such works are by definition referential" (FRBR p.67)

*Referential complement.* According to FRBR cadenzas, librettos, coreographies, and endings for unfinished works are defined as referential complements of works and expressions. Complements are intended to be integrated in some way with the other work. Anyway, they are not part of the original conception of that prior work, and sometimes they can be used on their own.

*Subject entities.* The "has as subject" relationship indicated that any of the entities in the model described in FRBR may be the subject of a work. This "has as subject" relationship could be extracted from special Marc fields concerning all kinds of subject headings (6XX) and classifications. According to FRBR the logical connection between a work and a related subject entity serves as the basis both for identifying the subject of an individual work and for ensuring that all works relevant to a given subject are linked to that subject.

*Navigating information.* For each work heading we also give the number



FRBR attribute	NORMARC	FinMARC	Selected
form of expression	leader, pos.6	leader, pos.6	no
language of expression	008, pos. 35-37	041 \$a	yes
other disting. characteristics			no
use restrictions		507	no
revision	250 \$a	250 \$a	no
translation			see text
referential successor			
referential supplement			
referential complement			

Table 7: *Attributes of high value in order to **select** an expression.*

of expressions related.

### Sorting the works

We could think of two logical ways of sorting the works. The first would be something like what James Duff Brown proposed: in chronological order of creation, but as neither the cataloguing rules nor the MARC fields give any information on this, we had to stick to the second to best: alphabetically.

In this project we had to do the sorting, not according to filing information found in indicators and sorting subfields, but straight alphabetically from the first letter. This is because in many cases we found the work titles in fields where no filing information was provided (500 fields, titles separated by semi colon in one single subfield, repeated \$a in 241 and so on).

#### 5.1.2 The expression

As the expression level is a new concept in cataloguing most attributes relevant for this entity level naturally are not present in the record in a form suitable for machine handling. Usually we find this kind of information in notes fields, some times it coincides with the same information on the manifestation level.

In table 7 we list the FRBR attributes of high value when it comes to selecting expressions. Discussion of the referential attributes is given above. Our selection of elements fits the actual type of material we are studying, but should be extended.

We found the language of expression to be the most prominent attribute and have chosen to present this first for each expression. Even though we

opened the language code we felt this was not enough for selecting. We would like to have the title of the expression also. This attribute is not present in the record, but as it is identical to the manifestation title in many cases, we felt free to use one of the manifestation titles for this purpose, see figures 2 and 5.

*Form of expression.* The form of expression is the means by which the work is realized. It is indicated e.g. on the leader as record type information, or as notes on nature, scope, literary form etc. (FRBR p. 36). This kind of information is usually given in the note fields.

*Other distinguishing characteristics.* According to FRBR a distinguishing characteristic is any characteristic of the expression that serves to differentiate the expression from another expression of the same work (e.g. an edition or version designation etc.).

*Use restrictions* are restrictions on access to and use of an expression. Usually the use restrictions are indicated in notes.

*Translation.* When we are able to identify an expression as a translation by combining information from different subfields, we present it after the original expression.

### Sorting the expressions

We did the sorting by language codes. With the codes opened, this does not appear 100% correct, e.g. scr = Croatian, but the intention is illustrated. In the records from the national bibliographies it is possible to decide whether or not the language code represents the original language. In this case we made a special fix so the expressions in the original language will appear first in the list with a special remark stating this fact (see figure 2 on page 32).

#### 5.1.3 The manifestation

By definition a new ISBN reflects a new manifestation, so this element is important for differentiating between the manifestations of the same expression. However, this number is naturally not among the ones important for the end user to select a specific manifestation (unless he is a book agent).

More important to the end user would be the *title of the manifestation*, even though this attribute is not among the ones regarded important by the FRBR study. The *statement of responsibility* usually holds information both on creator(s), illustrator(s) and translator(s). This information is not always found in other parts of the record, but is nevertheless very important for selecting and must be a part of the presentation. *Date of publication* also plays a role for many users who want to get either the first or the last edition of a specific expression.

*Form of carrier.* The form of carrier is usually given as specific material designation, accompanying material or as notes. We did not take information codes concerning the form of release, accompanying material etc. into

FRBR attribute	NORMARC	FinMARC	Selected
statement of responsibility	245 \$c	245 \$d \$e	yes
edition/issue designation	250 \$a	250 \$a	yes
date of publication	260 \$c	260 \$c	yes
form of carrier	245 \$h	300 \$z	yes
reduction ratio	?	?	no
presentation format	?	?	no
system requirements	500 \$a	548 \$a	no

Table 8: *Attributes of high value in order to **select** a manifestation.*

consideration, because our material consists mainly of textual material.

*Edition.* The edition/issue indicates differences between the manifestations of the same work or expression. The edition statement in the field 250 is useful for the user which is selecting the newest edition or version of the work. The edition statements in the field (555 in Finmarc) indicate that the editions described in the content note are equal.

*Reduction ratio.* We did not take this into consideration, because this attribute applies to the microforms only.

*Presentation format.* We did not take this into consideration, because this attribute applies to the visual projections only.

*System requirements.* We did not take this into consideration, because this attribute applies to the electronic resources only.

### Sorting the manifestations

There are several possibilities here, so a complete system should let the user select the sorting sequence himself. For some users the form of carrier is most important, for others the year of publication. The chosen sorting element(s) should get a prominent position in the presentation, so the sorting appears easily identifiable.

In our presentation we have sorted the manifestations by date of publication when they are shown in a separate window (see figure 3). In the special case of presenting the manifestations directly under the work heading the sorting is done by two elements, first by language (code) and second by date of publication (see figure 4).

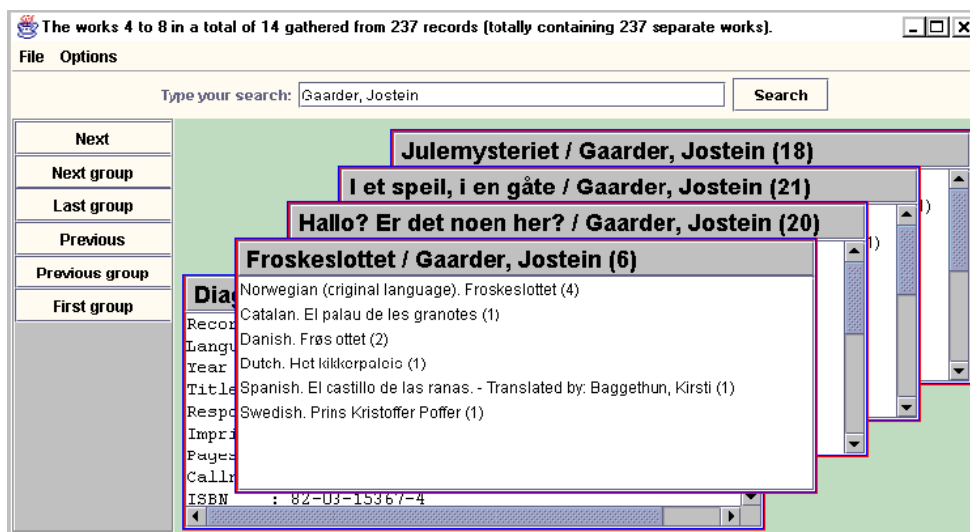


Figure 2: *Card catalogue 1*. The card headings are composed of work title and person responsible for the work. If there are many manifestations, we present only the expressions, here represented by document language and manifestation title (see text), original language on the top and then languages alphabetically. To each expression is added the number of manifestations of this expression.

## 5.2 The card catalogue idea

This idea is based on a sketch presented in a report on architecture for digital libraries[1]. The hit list is presented by overlapping cards. The top of cards may all be seen as one horizontal axis going into the screen of a two axis system and the card face as another – vertical – axis differentiating the information under the top heading of each card.

The two axis might be *authors – works*, *works – expressions*, *expressions – manifestations*, *subjects – works*. This would depend on the search performed. If the user searches for a specific author and there are several hits, the horizontal axis will list the persons names, and the vertical axis the works of each person. If, on the other hand, the search results in only one hit, the horizontal axis will contain the works of the author, and the vertical axis the various expressions.

In figure 2 this interface is shown with the works of *Jostein Gaarder* along the horizontal axis and the expressions along the vertical axis under each work. The number of expressions identified for each work is given in parenthesis at the end of the work heading. The number of manifestations for each expression is found at the end of the expression information (see figure 2).

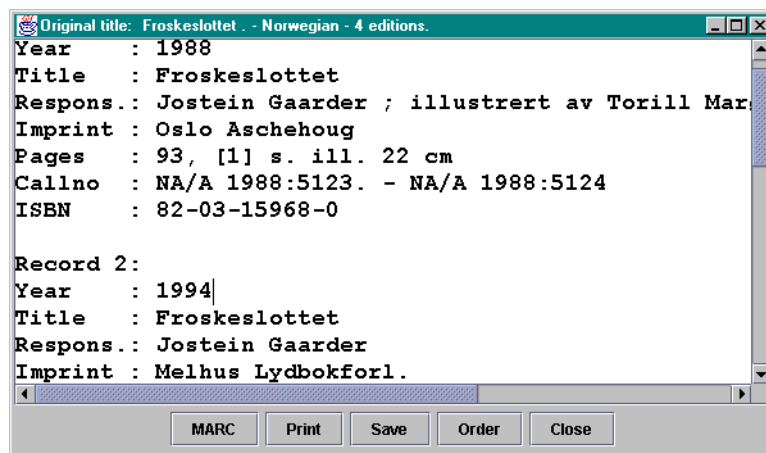


Figure 3: *The manifestations. The sorting is by publishing year.*

The user might select a specific expression thereby initiating a separate window containing all the manifestations under this expression (see figure 3). The manifestations of the chosen expression are sorted chronologically, and the publishing year is presented first for each record. The window contains buttons with various functions such as *print*, *save* and *order* and the window does not disappear unless the user explicitly closes it. This means that the users can keep as many sets of manifestations as they want, all in separate windows.

If a work appear in only a few manifestations (the number being user defined), they might be presented directly along the vertical axis(see figure 4). This saves the user from unnecessary clicking.

One might imagine that in a database where the data is in the structure of the FRBR model, the card might be expanded to hold links to related works (abbreviations, about this work and so on).

Please note that the frame header contains navigational information. The header states the size of the hit list and which work of the hit list is shown at the moment. The user might navigate back and forth through the hit list either work by work or by group of works. The number of cards shown might be defined by the user by a menu choice.

In the header of the manifestations window we find the original title, the current language (expression) and the number of manifestations (editions).

### 5.3 The tree structure idea

The second user interface presents the works as nodes or branches of a hit list in a tree-like structure (see figure 5). The works appear sorted alphabetically on original title and the number in the end of the title indicates how many

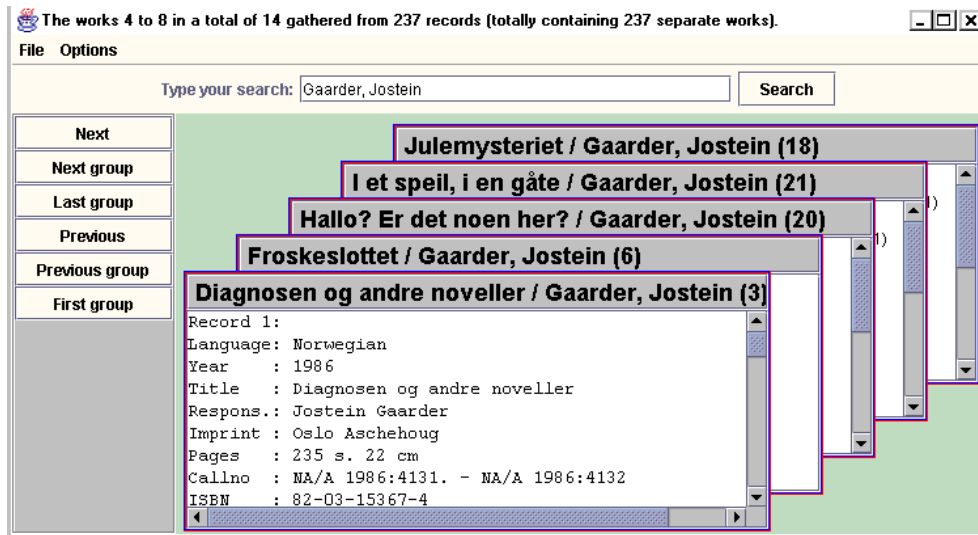


Figure 4: *Card catalogue 2.* If there are few manifestations under a work heading, they are presented directly on the card in alphabetical order according to language and within each language, chronologically.

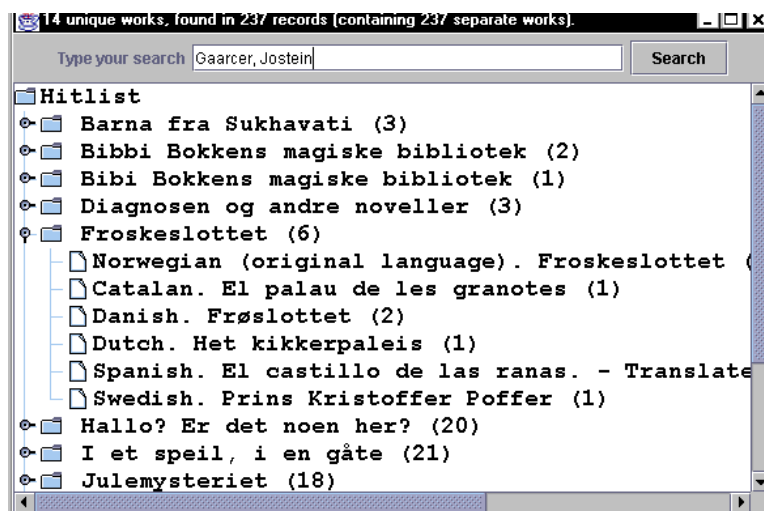


Figure 5: *Tree structure catalogue.* The work titles are sorted alphabetically and are expandable branches. The leaves represent expressions.

expressions are identified under this work.

The work nodes are expandable with the expressions as leaves. The expressions are sorted in the same way as for the interface previously described, original language on the top and then languages alphabetically.

These leaves are active, a click initiates a manifestations window as before (see figure 3).

The top node of the tree is here marked with the text *Hitlist*. In a well developed system this should of course be the search string and the tree could consist of several top nodes representing several hit lists.

## 6 Summary

The cataloguing rules reflect the technology of the card catalogue and are well suited for this purpose. Unfortunately, moving the data from the card catalogue to the MARC format kept the card characteristics in the data and in the beginning the purpose of the MARC records was chiefly to be a means of making the production and distribution of the cards more effective.

During the filing process in the card catalogue a structure was added that is usually not seen in the databases.

### **Functions in 700**

One of the main reasons for the noise we have experienced through our experiments stems from the unqualified use of the 700 field and the way the systems index this information together with the 100 field regardless of the real function the person has.

We feel that function in the 700 field should be mandatory in the cataloguing rules and that the list of functions should also be expanded. The systems must use function to present more structured hit lists for the end user.

This would make it possible for systems to present all functions a person might have in the database, making it possible for the end users to choose either all or some of the functions in their bibliographical navigation.

### **Work title (original titles)**

Our investigation also showed that it would be an advantage for the analysis if original titles could be entered in a more consistent way, in separate, repeatable fields, and this must also be done even when the manifestation title and the work title are the same (especially relevant when there are several works within the same manifestation).

### **Language codes and language of original work**

Language codes are one of the most important attributes to identify the different expressions of a specific work. It is also important in order to

identify whether a manifestation is a translation or not.

### **Authority files**

Normalization of data is a linking device in itself. Using authority files in the bibliographic environment helps to establish the navigational structure, both by collocating entities and differentiating between them. The most common authority files used are the names of persons and corporate bodies. Our work shows that there is also a need for work authorities to be able to collocate the same work under one heading.

### **The work done by cataloguers**

Cataloguing takes time, it is labour intensive work. This has misled many people to believe that cataloguing must be simplified and these people have come up with various suggestions that diminish the quality of the catalogue. One such example is a decision to leave out the function from the 700 field.

We believe to have shown that the work of cataloguers (who understand the principles and logic of the rules) must be given high priority and that the structure established by using the rules saves a lot of time for the end user. And as this is done for every user every day now and in the future the total time saved will be enormous if we today put the extra work into cataloguing properly.

### **Elaborating the user interface, other ideas**

We have shown that it is possible to put more effort in presenting the hit lists of the OPACs by using the structure hidden in the bibliographic data.

We have in this report presented two simple ideas for OPAC user interfaces. These interfaces only let us navigate between the entities work, expressions and manifestations. Using more of the information that is actually present in the records, it would be easy to establish links from works to other works (e.g. works *about* a certain work using relational fields in the 6XX and 7XX fields), from creators to works *about* this creator and so on.

To get inspiration for library catalogue interfaces one should also look at what kinds of graphical tools are used in other computer applications, and see how these tools might improve the presentation of the bibliographical structure that we put into our catalogues.

## **7 Testing the methods on BIBSYS data**

BIBSYS is a governmentally financed library system serving most university and college libraries in Norway. The database contains 2,5 million records and around 8 million holdings. The cataloguing holds different degrees of quality called *kat2*, *ukatm* and *konv*<sup>15</sup>. The first category is expected to be

---

<sup>15</sup>There is also a category called *ukat* for items still in the acquisition process.



Author	Number of records	Number of work lines	Number of unique work lines
Wassmo, Herbjørg	171	176	33
Ibsen, Henrik	2 643	2 994	584
Gaarder, Jostein	301	299	45
Solstad, Dag	96	98	43

Table 9: *Results of using the work reduction procedure on records from the BIBSYS catalogue.*

high quality catalogued items with respect to the cataloguing rules, but the last two are in most cases minimal catalogued items (converted data from card catalogues, cataloguing done by non-cataloguers and so on). About 53% of the records are marked as of high quality. There is no way to restrict a search to items of a distinct item quality. As will be seen, this is reflected in the results.

The BIBSYS system has a special logic for dealing with *multipart items*. This called for a reprogramming when dealing with BIBSYS records. Somewhat simplified one might say that some information common for all the parts resides in a record on a higher level, while information special for the single part is entered in the part record. To do a proper analysis we first had to make special search for the main records and include these in the search set. This was a cumbersome process, but it would be easy to automate this in an online environment. Second, when dealing with BIBSYS records the analyzing program was revised to copy information from the high level record to the part record, when these data were not present in the part record. This copying included information like author, language code, publisher, place of publication and so on.

In the BIBYS database we find original titles in field 241 and with a qualifying prefix in the 500 note field (not both at the same time). This reflects a change of rules over time. The 241 field is sometimes repeated for each work title, sometimes several work titles appear as repeated subfield \$a within one occurrence of 241. Usually, in field 500, the work titles are separated by *space-semicolon-space*, but not always. Thanks to Anne Munkebyaune (BIBSYS) who answered questions about cataloguing rules and practice in the BIBSYS catalogue.

## 7.1 Herbjørg Wassmo

We start this analysis with the Norwegian author *Herbjørg Wassmo* as a tribute to *Ella Leth-Olsen* who works in the BIBSYS organization and who in a lunch some years ago inspired Knut Hegna to go into the problems of today's online catalogues by presenting to him the *Herbjørg Wassmo problem*

(HWP).

As a cataloguer Ella was frustrated because even though the author had only made 17 works (at that time), the hit list in BIBSYS was made up of an unordered list of 150 items with no structure at all. Ella felt that much could be done if only the bibliographic information in the records was used properly.

At the time of the analysis (November 2001) searching for *Herbjørg Wassmo* resulted in 171 records. Our algorithms found 176 works in these records, which was reduced to 33 unique works.

By looking through the work titles we see that 12 works were due to lack of original title in the records. We have not checked this, but the reason might be that these records belong to the low quality records of the catalogue. One record represented bad cataloguing and one record was *about* one of the short stories, so Wassmo had been entered in the 700 field.

This 700 field also contained the title of the short story which points at leaving this field out of the analysis in the next version of the program. It might be taken into consideration when dealing with relations to the specific work.

All in all, when subtracting these 14 records from the total of 33 we reach 19 real works by this author.

In the BIBSYS records it was not possible to decide whether an expression is in the original language or not, because original language is not entered in the 041 field. One might expect that the absence of original title (field 241) would be strong indication of an expression in the original language, but as we have seen we can't trust the records in this respect. For this reason we could not put expressions in the original language on the top of list.

It is also difficult to identify the translator as this information is mainly entered in the *statement of responsibility*, 245 \$c, and it would be hazardous to pick the first name from the 700 field. For this reason, we could not distinguish between different translations to the same language.

## 7.2 Henrik Ibsen

Searching for *Henrik Ibsen* in the BIBSYS catalogue results in a hit list of 2643 records. In the OPACs of BIBSYS only 1000 might be shown and the records come in no order at all. The sequence changes also from day to day. Clearly, something has to be done.

In these 2643 records we found 2994 works. When using the work reduction algorithm the list of works is reduced to 584. Even this is too much. By reading through one third of the records, we found that 32 were works *about* Ibsen or his works; 48 records were actually musical scores with text by Ibsen; 38 were programs for theater performances with Ibsen as added entry in the 700 field; 30 records missing original title; four parodies; 16

adaptations. We also found 10 parts of his works (act 1 and act 2 and so on of a play in separate manifestations) with no obvious logic to collocate these under the same work heading.

Extrapolating these findings to the whole set and subtracting these from the 584 we reach a set of about 130 works *by* Ibsen.

A closer study shows that among the 584 unique works we could easily identify 77 works that were obviously created by other persons than Ibsen. The work distribution statistics<sup>16</sup> showed that 465 works appeared in only one expression and one manifestation. These represent a strong indication of works not created by Ibsen, missing original titles, about Ibsen, adaptations and so on. The reason the records were present in the hit list can be found in the unqualified use of the 700 field and the system's indexing of this field together with the 100 field.

As translators were not identifiable we could not distinguish between different translations into the same language as separate expressions.

### 7.3 Jostein Gaarder

The BIBSYS catalogue contained 299 records by this author<sup>17</sup>. We were able to identify 45 unique works among these. In the Norwegian national bibliography the number of unique works was found to be 13. It seems a lot of works where Gaarder is co-author are not included in the national bibliography. We have not had time to check this. The reason might be that these are philosophy text books for schools which are not of high priority in the national bibliography, but these works are important in libraries at the colleges educating teachers.

Among the 45 works we found eight extras due to missing registration of original titles and six with typos in the title.

### 7.4 Dag Solstad

The hit list contained 96 records and the program found 98 works within this set. The number of unique works was 43. The analysis of national bibliography showed 32 unique works.

First, one of the extra works was caused by a misplaced semicolon in a title, thereby leading to the interpretation of two titles instead of one. Next, two works were victims of bad cataloguing<sup>18</sup>. Three records were missing original titles giving cause to three separate works. One work was caused by a typo in the title information. This brings the number of works down to 36. This catalogue included also some works by Dag Solstad that are missed

---

<sup>16</sup>See the same kind of statistics for Gaarder in table 5.

<sup>17</sup>See section 3.2 for a short biography of Jostein Gaarder

<sup>18</sup>Probably these were of the low quality type. See the introduction to section 7.

by the national bibliography (analytic entries, theater program texts and so on).

The analysis of the expressions showed that errors were mostly caused by missing language codes in (14 records probably in the low quality category).

## 7.5 BIBSYS catalogues: summary

### Several catalogues and cataloguing levels

Our analysis is clearly suffering from the different cataloguing practices which exist inside the database, both from changing rules and formats over the years and also from differences formed by the merging of cataloguing data from many different libraries during the last 20 years. Authority files for personal names and corporate bodies have been introduced some years ago, but as they are based on the data already entered in the database it will take some time before they reach full effect in a database consisting of 2.5 million records.

### The use of the 700 field and the missing function

As we have seen, the lack of function in the 700 field also leads to a lot of noise in the hit lists when searching for a specific person in the function of creator. This noise appears as a lot of unique works in our analysis. If function had been used and the search system allowed to specify which function we were looking for, the analysis would have turned out better. Very often, when a work is *about* a person an entry is correctly made in the 600 field, but this entry is also supplemented with an entry in the 700 field without any function accompanying it.

### Language codes and translators

Lastly, we must point to the missing language codes in many records which make it impossible to differentiate between the various expressions. There is no way to establish whether a certain manifestation is a translation or not.

## 8 Moving on

The FRBR study has already had great impact, both seen from the cataloguer's point of view and as a way of explaining the bibliographical structure to the people making the systems. The study has functioned like a rosetta stone in the communication between the cataloguers and the computer people and by this, has created a lot of enthusiasm and discussions around cataloguing questions, maybe as much as at the time of Charles A. Cutter who declared in 1904:

*Still I can not help thinking  
that the golden age of cataloging is over,  
and that the difficulties and discussions  
which have furnished an innocent pleasure to so many  
will interest them no more.  
Another lost art.[3]*

This gives hope for the future to bring the same balance between rules and technology that existed in the times of the card catalogue, to the computer and database environment.

We will here mention some initiatives that we find important in the further discussions and development.

#### **Tom Delsey's MARC/FRBR/AACR analysis**

The present study has a practical starting point: what kind of structures and problems do we find in the existing records. Tom Delsey's[4] thorough study is done from a more theoretical point of view and covers the whole field of connections between the cataloguing rules (AACR), the data transport format (MARC) and the entities, relations and attributes of the FRBR model.

#### **JSCs work on expression level cataloguing.**

The expression is a new concept in cataloguing and consequently the cataloguing rules at the moment do not meet the sufficient requirements to identify this level. One step towards establishing this comes from JSC format variation working group[6].

#### **The FRANAR working group**

Interoperability is not only a technical and format matter. A lot of emphasis must be put on the data residing inside the record. Real national and international interoperability acquires the common use rules and authority data. Doing without authorized data would mean a lot of duplicates and noise in a multicasting situation, e.g. using the Z39.50 protocols to create virtual union catalogues.

The FRANAR working group lays the foundation for international cooperation on authority data, thereby moving the library world towards the ultimate objective: Universal Bibliographic Control.

#### **The META-LIB project in Germany**

"The project Meta-Lib, sponsored by the German Research Foundation, analyses the increasing importance of metadata for libraries and learned societies. The project partners Die Deutsche Bibliothek and SUB Göttingen develop guidelines for the use of metadata based on the DCMES, in the in-

dexing of electronic documents. In addition, they provide information on the discussion about metadata and on various metadata formats.”(From the project presentation).

### **Relations**

One point we did not have time to look into is *relations* other than the ones we tried to identify between expressions and manifestations of the same work. Clearly we could have made more out of the relations established through the use of data from the 6XX and 7XX fields. The problem with these kinds of relations is that they in most cases are not real links to other records or works, only back to the record or work in hand.

We recommend for further study Bernhard Eversbergs[5] suggestions for the use 787 field as a simple linking field. His proposals would be easy to implement in most of today’s systems and meet a lot of the relational requirements of the FRBR study.

## **A Project description**

### **AGREEMENT**

Agreement between BIBSYS and Helsinki University Library (HUL) concerning models and practical applications of the IFLA study on Functional Requirements for Bibliographic Records, (FRBR).

### **Working Group**

Knut Hegna and Eeva Murtomaa with cataloguing experts supporting them. If, for any reason, one of the persons involved can no longer participate, their institution is not committed to the project.

### **Duration**

The project will start March 15th 2001 and end in February 28th 2002.

### **Costs**

BIBSYS and HUL will cover the project costs from their basic budget. The participants will make their own agreements with their institution in this regard.

### **Goals**

#### **1. Analyzing hit lists**

a) Investigating the possibility to construct an algorithm for analyzing a set of hits from an OPAC search in the light of FRBR. That is, trying to identify works, expressions, and manifestations on the basis of the MARC data of the records in the hit list.

b) Identify problem areas in this regard.

The investigation starts with hit lists based on author search and then moves on to title and subject search.

The algorithm must be able to cope with different MARC formats.

The test data of the project is gathered from both the Finnish national bibliography and from the Norwegian national bibliography

#### **2. Identify a work level**

Based on the results from step 1) try to recognize the problems which arise when trying to identify a work level in the catalogue as a whole.

#### **3. OPAC design**

Using the possible results from steps 1) and 2) and on the basis of IFLAs 'Guidelines for OPAC displays' try to design a new kind of user interface. The user interface should be multilingual (English, Finnish and Norwegian) and handle different MARC formats (MARC21, FIN-MARC, NORMARC).

#### 4. Testing with other data sets

Depending on the results of 1) and 2) the methods will be tested with data sets from the BIBSYS catalogue in order to identify new problem areas.

The findings of the project will be published in a final report.  
Roy Gundersen (BIBSYS), Esko Häkli (HUL)

## B Example records: single works by Tove Jansson

Title of the work:

```
*041 $a swe
*080 $a 839.79-3(024.7)
*1001 $a Jansson $h Tove
*245 $a Trollvinter $d Tove Jansson
      $e teckningar och omslag av författarinnan
```

Expression of the same work:

```
*041 $a eng $c swe
*080 $a 839.79-3
*1001 $a Jansson $h Tove
*241 $a Trollvinter
*2452 $a Moominland midwinter $d written and ill. by Tove Jansson
      $e transl. by Thomas Warburton
*70011$a Warburton $h Thomas
```

Translation (expression) based on another translation (expression) of the work:

```
*0411 $a jpn $c swe
*080 $a 839.7-3(024.7)
*1001 $a Jansson $h Tove
*241 $a Trollkarlens hatt
*2452 $a Tanoshii Moomin ikka $d Tove Jansson
      $e [käänt.] Yamamuro Shizuka
*500 $a Käänetty engl. painoksesta Finn family Moomintroll
```

## C Example records: several works in a single manifestation

### C.1 Norway

These examples<sup>19</sup> are the basis of the rules governing the separation of different works from a single manifestation.

The examples are records found when searching for *Ibsen, Henrik*<sup>20</sup>.

---

<sup>19</sup>The examples are somewhat edited so the points are emphasized.

<sup>20</sup>The fields referred to: 245 - title proper; 500 - note; 505 - content note; 700 - added entry personal name; 740 - added entry title; 960 - local extension, work authority.



The first example is a 1978 record containing a Polish edition of five plays. No Polish expression titles are given - the manifestation title is *Dramaty* (dramas). The Norwegian work titles are given in the note field 500 prefixed with the text *Inneholder:* (contains) and the titles separated from each other with semicolons.

```
*24500$a Dramaty
*50000$a Inneholder: Et dukkehjem ; Gengangere ; En Folkefiende ;
Vildanden ; Hedda Gabler
```

The next example is a 1978 record containing a manifestation of three works in Lithuanian. Only original titles are given in the 500 field (this time with the prefix *Innhold:* (Content)).

```
*24500$a Dramos
*50000$a Innhold: Hærmændene på Helgeland ; Hedda Gabler ;
Bygmester Solness ; John Gabriel Borkman
```

This next example is also from 1978 and contains two plays in Dutch. Manifestation titles in 245\$a separated with a semicolon. Original titles in field 500.

```
*24500$a De wilde eend ; Hedda Gabler
*50000$a Originaltitler: Vildanden ; Hedda Gabler
*96000$a Ibsen, Henrik $d (1828-1906) $t Hedda Gabler
```

In the following example from 1982 the expression titles of the five plays are given in the 245\$b subfield and the original titles in the 245\$t subfield (undefined in NORMARC as of 2001). All titles are separated with semicolons and follow the same sequence.

```
*24500$a Drami
$b Stfulbovete na obistestvoto ; Kuklen dom ; Prizraci ;
Naroden vrag ; Kogato nie, mfurtvite se sfubudim
$t Samfunnets støtter ; Et dukkehjem ; Gengangere ;
En folkefiende ; Når vi døde vågner
```

The next example is a 1987 record containing a Croatian edition of three plays with *Drame* as the manifestation title. The separate work titles are given in the note field 505 beginning with a volume number and the titles in Croatian separated by semicolons. The Norwegian original titles follow after the prefix *Originaltitler:*. The note follows the ISBD rules for delimiting characters.

```
*2450 $a Drame
*505 $a 2 : Rosmersholm ; Heda Gabler ; Jon Gabrijel Borkman /
[prevod Zeina Mehmedbaisibc]. - 1978. - 348, [2] s. -
Originaltitler: Rosmersholm ; Hedda Gabler ;
John Gabriel Borkman. -
Originalutgave: København : Gyldendal, henholdsvis 1886,
1890, 1896. -
Har bibliografi
```

What follows is a two volume record from 1987 (two manifestations in one record). The content of each volume is given in two 505 note fields. The expression titles prefixed with the text *Innhold:* (Content) and the original titles prefixed with the text *Originaltitler:* (originaltitler). The ISBD character delimiters are not 100%.

```
*24510$a Dramebi
*505 $a T. 1. - 1994. - 590 s. - Innhold: Keiser da galileveli ;
      Sazogadoebis burjni ; T'ojinebis saxli ; Mevcwvenebani ;
      Xalxis mteri.
      Originaltitler: Keiser og Galilæer ; Samfundets støtter ;
      Et dukkehjem ; Gengangere ; En folkefiende
*505 $a T. 2. - 1995. - 542 s. - Innhold: Gareuli ixvi ;
      Rosmersholmi ; Zggvis asuri ; Heda Gableri ; Mvsenebeli
      Solnesi ; Datara Eioldp'i ; Roc'a vcwven, mkvdrebi,
      vigvizebt.
      Originaltitler: Vildanden ; Rosmersholm ; Fruen fra havet ;
      Hedda Gabler ; Bygmester Solness ; Lille Eyolf ;
      Naar vi døde vågner
```

The next record, of unknown date, includes three plays in English. Expression titles both in subfield 245\$b and as title added entry in field 740. Original titles are given in the 240 field as repeated subfield \$a which violates the rules of the format.

```
*2401 $a En folkefiende $a Vildanden $a Hedda Gabler
*2451 $a Three plays
      $b An enemy of the people ; The wild duck ; Hedda Gabler
*740 $a An enemy of the people
*740 $a The wild duck
*740 $a Hedda Gabler
```

A Norwegian 1989 edition with three work titles all in the 245-field separated with semicolons. The three titles are also given in the 740 field (title added entry). Finally, two of the three titles are found as work authority headers in field 960.

```
*24503$a Et dukkehjem ; Gengangere ; Vildanden $c Henrik Ibsen
*74000$a Et dukkehjem $w 3
*74000$a Gengangere
*74000$a Vildanden
*96000$a Ibsen, Henrik $d 1828-1906 $t Gengangere
*96000$a Ibsen, Henrik $d 1828-1906 $t Vildanden
```

Three records from a four volume manifestation in French added in 1992/3 including nine plays altogether.

```
*24504$a Les douze dernières pièces
*505 $a Vol. 1 : Les piliers de la société ; Maison de poupée ;
      Les revenants. - 1990. - 370, [5] s. -
      Originaltitler: Samfundets støtter ; Et dukkehjem ; Gengangere
*700 2$a Ibsen, Henrik $d 1828-1906 $t Les piliers de la société $u 4
*700 2$a Ibsen, Henrik $d 1828-1906 $t Maison de poupée
```

\*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t Les revenants \$u 4  
 \*7400 \$a Samfundets støtter  
 \*7403 \$a Et dukkehjem  
 \*7400 \$a Gengangere  
 -----  
 \*505 \$a Vol. 2 : Un ennemi du peuple ; Le canard sauvage ;  
       Rosmersholm / [Rosmersholm] traduit par Terje Sinding et  
       Bernard Dort. - 1991. - 403 s. -  
       Originaltitler: En folkefiende ; Vildanden ; Rosmersholm  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t Un ennemi du peuple \$u 3  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t Le canard sauvage \$u 3  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t Rosmersholm  
 \*7403 \$a En folkefiende  
 \*7400 \$a Vildanden  
 -----  
 \*505 \$a Vol. 4 : Le petit Eyolf ; John Gabriel Borkman ;  
       Quand nous nous réveillerons d'entre les morts. -  
       1993. - 296 s. -  
       Originaltitler: Lille Eyolf ; John Gabriel Borkman ;  
       Når vi døde vågner  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t Le petit Eyolf \$u 3  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t John Gabriel Borkman  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t Quand nous nous réveillerons  
       d'entre les morts  
 \*7400 \$a Lille Eyolf

In this 1993 record, English titles are given in the 245 field and the last two also as authority headers in 700. Original titles are given both in field 500 prefixed by *Originaltitler* and in the 740 fields. For some reason, *A doll's house* also appears in the 740 field, which makes it impossible to distinguish it from the original titles.

\*24502\$a A doll's house ; The wild duck ; The lady from the sea  
 \*500 \$a Originaltitler: Et dukkehjem ; Vildanden ; Fruen fra havet. -  
       1. engelske utgaver Copenhagen 1880, London 1897 og London 1890  
 \*74020\$a A doll's house  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t The wild duck \$u 4  
 \*700 2\$a Ibsen, Henrik \$d 1828-1906 \$t The lady from the sea \$u 4  
 \*7403 \$a Et dukkehjem  
 \*740 \$a Vildanden  
 \*740 \$a Fruen fra havet

In these two last examples one may see a shift towards a situation with data more suitable for machine handling, but there is still some way to go.

## C.2 Finland

### Collection without a collective title.

Separate works/expression titles given on the title and statement of the responsibility area.

\*021 \$a 91-0-056156-8 \$c sid.  
 \*041 \$a swe  
 \*080 \$a 839.79-3  
 \*080 \$a 929 Jansson  
 \*080 \$a 82  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Bildhuggarens dotter och Sommarboken \$d Tove Jansson  
 \*260 \$a [Stockholm] \$b Bonnier \$c 1996 \$f WSOY  
 \*300 \$a [322] s. \$b kuv. \$c 22 cm  
 \*60011\$a Jansson\$h Tove  
 \*745 \$a Sommarboken

\*1001 \$a Jansson \$h Tove  
 \*241 \$a Kometjakten  
 \*241 \$a Trollkarlens hatt  
 \*2452 \$a Muumipeikko ja pyrstötähti  
     \$a Taikurin hattu \$d Tove Jansson  
     \$e suomentanut Laila Järvinen \$e [tekijän kuvittama]  
 \*260 \$a Porvoo \$a Helsinki \$a Juva \$b WSOY \$c 1993  
 \*300 \$a 155, 147, [2] s. \$b kuv. \$c 18 cm  
 \*440 \$a Kuukauden pökkari  
 \*70011\$a Järvinen \$h Laila  
 \*745 \$a Taikurin hattu

### Collections with a collective title.

Separate work/expression titles given as content and original title notes.

\*1001 \$a Jansson \$h Tove  
 \*2452 \$a Muumilaakson tarinoita  
 \*505 \$a Yhteislaitos... teoksista: Trollkarlens hatt, Taikurin hattu ;  
     Farlig midsommar, Vaarallinen juhannus ; Trollvinter,  
     Taikatalvi  
 \*745 \$a Taikurin hattu  
 \*745 \$a Vaarallinen juhannus  
 \*745 \$a Taikatalvi

\*100 \$a Jansson \$h Tove  
 \*241 \$a Trollvinter  
 \*241 \$a Trollkarlens hatt  
 \*241 \$a Muminpappas memoarer  
 \*241 \$a Pappan och havet  
 \*241 4\$a Det osynliga barnet  
 \*2452 \$a Muumitroll \$d Tove Jansson \$e [autori illustratsioonid]  
 \*260 \$a Tallinn \$b Tiritamm \$c 1994  
 \*300 \$a 523, [3] s. \$b kuv. \$c 25 cm  
 \*505 \$a Sisältö: Volukubar ; Trollitalv ; Muumipapa memuaarid ;  
     Nähtamatu laps ja teisi jutte ; Muumipapa ja meri  
 \*745 \$a Volukubar  
 \*745 \$a Trollitalv  
 \*745 \$a Muumipapa memuaarid  
 \*745 \$a Nähtamatu laps ja teisi jutte  
 \*745 \$a Muumipapa ja meri

\*1001 \$a Jansson \$h Tove  
 \*2452 \$a Muumiperheen tarina \$d Tove Jansson \$e kuv. Tove Jansson  
 \*260 \$a Porvoo \$a Hki \$a Juva \$b WSOY \$c 1987 \$e Porvoo  
 \*300 \$a 164, [1] s. \$b kuv. \$c 29 cm  
 \*505 \$a Sisältö: Taikatalvi / suom. Laila Järvinen. -  
     Alkuteos: Trollvinter. - 8. p. Muumipappa ja meri /  
     suom. Laila Järvinen. - Alkuteos: Pappan och havet. -  
     7. p. Muumilaakson marraskuu / suom. Kaarina  
     Helakisa. - Alkuteos: Sent i november. - 8. p  
 \*555 \$a 2. p. 1988. - 3. p., [12. tuhat] 1989. - 4. p.,  
     [15. tuhat] 1991. - 5. p., [20. tuhat] 1991. - 6. p.,  
     23. tuhat 1992. - 7. p., 25. tuhat 1993. - 8. p.,  
     27. tuhat 1994  
 \*745 \$a Taikatalvi  
 \*745 \$a Muumipappa ja meri  
 \*745 \$a Muumilaakson marraskuu  
  
 \*1001 \$a Jansson \$h Tove  
 \*245 2\$a Takaisin Muumilaaksoon \$d Tove Jansson \$e kuv. Tove Jansson  
 \*260 \$a Porvoo \$a Hki \$a Juva \$b WSO Y \$c 1988 \$e Porvoo  
 \*250 \$a, [1] s. \$b kuv. \$ c 30 cm  
 \*500 \$a Alkuteokset: Kometjakten ; Trollkarlens hatt ;  
     Muminpappans bravader ; Farlig midsommar ;  
     Det osynliga barnet och andra berättelser  
 \*505 \$a Sisältö: Muumipeikko ja pyrstötähti ; Taikurin hattu ;  
     Muumipapan villi nuoruus ; Vaarallinen juhannus ;  
     Näkymätön lapsi / suom. Laila Järvinen  
 70011 \$a Järvinen \$h Laila

#### Separate titles connected to the ISBN number

\*021 \$a 951-0-24658-1 \$c koko teos  
 \*021 \$a 951-0-92609-4 \$c Muumilaakson kevätlaulu  
 \*021 \$a 951-0-92811-9 \$c Muumilaaksossa tapahtui  
 \*021 \$a 951-0-92610-8 \$c Muumiperheen pikku vieraat  
 \*0411 \$a fin \$c swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Muumilaakso \$z Aänite \$d Tove Jansson  
     \$e suomennos: Laila Järvinen  
 \*260 \$a [Helsinki] \$b WSOY \$c 2000 \$f Kasettilinja  
 \*300 \$z 3 C-kas. \$b Dolby HX PRO  
 \*511 \$a Lukijat: Ahti Jokinen ja Tytti Paavolainen  
 \*70011\$a Järvinen \$h Laila  
 \*70011\$a Jokinen \$h Ahti  
 \*70011\$a Paavolainen \$h Tytti  
 \*745 \$a Muumilaakson kevätlaulu  
 \*745 \$a Muumilaaksossa tapahtui  
 \*745 \$a Muumiperheen pikku vieraat

**Collection of works/expressions as a part of a multivolume work.**  
 Separate work/expression titles are given as content and original title notes

\*021 1\$a 91-1-895331-X \$c nid.  
 \*041 \$a swe  
 \*080 \$a 741.5  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Mumintrollet \$d av Tove Jansson  
 \*248 \$g 4  
 \*260 \$a [Stockholm] \$b Norstedt \*260 1\$c 1989  
 \*300 1\$a 119 s. \$b kuv. \$c 18 x 23 cm  
 \*505 1\$a Innehåller episoderna: Mumintrollet blir förälskat ;  
           Mumintrollet och marsinnevånarna ; Muminfamiljen och havet  
 \*745 \$a Mumintrollet blir förälskat  
 \*745 \$a Mumintrollet och marsinnevånarna  
 \*745 \$a Muminfamiljen och havet

\*021 1\$a 951-0-16254-X \$c nid.  
 \*0411 \$a fin \$c swe  
 \*080 \$a 741.5  
 \*1001 \$a Jansson \$h Tove  
 \*2411 \$a Mumintrollet 4  
 \*2452 \$a Muumipeikko \$d Tove Jansson  
 \*248 \$g 4 \$d [suomennos: Juhani Tolvanen ja Anita Salmivuori] \$y 0004  
 \*260 \$a Porvoo \$a Helsinki \$a Juva \$b WSOY  
 \*260 1\$c 1990 \$e Juva  
 \*300 1\$a 119 s \$b kuv. \$c 18 x 21 cm  
 \*505 1\$a Sisältää jaksot: Muumipeikko rakastuu ; Muumipeikko ja  
           marsilaiset ; Muumiperhe ja meri  
 \*70011\$a Tolvanen \$h Juhani  
 \*70011\$a Salmivuori \$h Anita

## D Example records: several works with different authors

These examples are collected from the Norwegian national bibliography.

The first example is a record from 1992 with two plays, one by Ibsen (Ghosts) and another by Strindberg (Miss Julia). English uniform title is given in the 700 field for one of the plays, and the original titles are given both in 500 and 740. The ones given in 500 are separated by a dot not according to the rules. Even though all the information we need is there, we see no way to match authors, titles and original titles.

\*100 0\$a Ibsen, Henrik\$d1828-1906  
 \*24510\$a Ghosts \$b a family drama in three acts  
       \$c Henrik Ibsen. And Miss Julia : a naturalistic tragedy  
       / August Strindberg ; [both] translated and edited  
       by Thaddeus L. Torp  
 \*500 \$a Originaltitler: Gengangere. Fr#hoken Julie. - Har  
       bibliografi  
 \*700 2\$a Strindberg, August \$t Miss Julia  
 \*7400 \$a Gengangere  
 \*7400 \$a Fr#hoken Julie

The next example is a manifestation published in 1998, but was entered in the database in 2001. The manifestation contains three plays by three different authors. Manifestation titles, authors and translators are all found in the 500 field. Here also, the original titles are separated by a dot. In the 700 field we find both original titles and main titles, but we have no way to tell which is which.

```
*24510$a Sydney Theatre Company $b "word for word) $c edited by John
    Senczuk
*260 $a Wollongong, N.S.W. $b Five Islands Press$c1998
*300 $a 241 s.
*440 $a Australian playwright signature series
*500 $a Undertittel på omslaget og smusstittelbladet:
    Contemporary Australian translations
*500 $a Innhold: Medea / by Euripides ; translated by Gregory
    McCart. Hedda Gabler / by Henrik Ibsen ; translated by
    May-Brit Akerholt. Dance of death / by August Strindberg ;
    translated by May-Brit Akerholt. Originaltitler: Medea.
    Hedda Gabler. Dödsdansen
*700 1$a Senczuk, John $e red.
*700 0$a Euripides $t Medea
*700 0$a Ibsen, Henrik $t Hedda Gabler
*700 0$a Strindberg, August $t Dödsdansen
*700 0$a Strindberg, August $t Dance of death
*700 1$a Akerholt, May-Brit $e overs.
*991 $a Akerholt, May-Brit
```

The next example record covers 5 volumes of a series entitled “The national literature of Norway”. Titles and authors are given both in the note field (505) and the added entry field (700), but we cannot identify original titles from this information as the titles in 505 are not prefixed by *original title* and the 700 field contains no qualifier telling what kind of title is given.

```
*24510$a Norges nasjonallitteratur
*505 $a 26 : Alle mine dikt / Rolf Jacobsen. -
    340 s. - 1. utg. 1990
*505 $a 27 : Dikt i samling / Olav H. Hauge. -
    298 s. - 1. utg. Oslo : Noregs boklag, 1972
*505 $a 28 : Fyr og flamme / Kjartan Fløgstad. -
    407 s. - 1. utg. Oslo : Samlaget, 1980
*505 $a 29 : Roman 1987 / Dag Solstad. -
    341 s. - 1. utg. Oslo : Oktober, 1987
*505 $a 30 : Dinas bok / Herbjørg Wassmo. - 1996. -
    498 s. - 1. utg. 1989
*70010$a Jacobsen, Rolf
    $d 1907-1994
    $t Alle mine dikt
*70010$a Hauge, Olav H.
    $d 1908-1994
    $t Dikt i samling
*70010$a Fløgstad, Kjartan
    $d 1944-
    $t Fyr og flamme
```

\*70010\$a Solstad, Dag  
 \$d 1941-  
 \$t Roman 1987  
 \*70010\$a Wassmo, Herbjørg  
 \$d 1942-  
 \$t Dinas bok

## E Example records: hidden relationships

Relationship: Equivalent with the original work: exact reproduction (facsimile).

\*041 \$a swe  
 \*080 \$a 839.79-3  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Karin, min vän \$d Tove Jansson  
 \*260 \$a [Stockholm] \$b Bonnier \$c 1987  
 \*300 \$a 26, [3] s. \$c 30 cm  
 \*500 \$a Texten är återgiven i faksimil efter manuskriptet

Same work/expression embodied in two manifestations. Relationship: derivative.

\*041 \$a swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Mumintrollet på kometjakt  
 \$d ritad och berättad av Tove Jansson  
 \*260 \$a Norrköping \$b Sörlins förlag \$c 1956  
 \*300 \$a 156 s. \$b kuv.

\*041 \$a swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Mumintrollet på kometjakt  
 \$d ritad och berättad av Tove Jansson  
 \*260 \$a Helsingfors \$b Söderström \$c 1956  
 \*300 \$a 156 s. \$b kuv.

Editions of the same work/expression. Relationship: derivative

\*1001 \$a Jansson \$h Tove  
 \*2452 \$a Takaisin Muumilaaksoon \$d Tove Jansson  
 \$e kuv. Tove Jansson  
 \*260 \$a Porvoo \$a Hki \$a Juva \$b WSOY \$c 1988 \$e Porvoo  
 \*300 \$a 250, [1] s. \$b kuv. \$c 30 cm  
 \*555 \$a 2. p. 1991. - 3. p. 1992. - 4. p. 1994

Derivative relationship: abridged and revised edition of the work (according to note ) and simultaneous publication (different ISBD-numbers indicating different publishers)



\*021 \$a 91-20-07415-8 \$c sid.  
 \*021 \$a 91-21-05804-0 \$c Almqvist & Wiksell, nid.  
 \*041 \$a swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*245 \$a Vinter i Mumindalen \$d Tove Jansson  
 \*260 \$a Stockholm \$b Awe/Gebers \$c 1985  
 \*300 \$a 45, [1] s. \$b kuv. \$c 19 cm  
 \*500 \$a Texten är tagen från Trollvinter (1957):  
     andra kapitlet (något rev.), tredje kapitlet (i sin helhet),  
     sjätte kapitlet (ett kort avsnitt)

Work: Jansson, Tove, Resa med lätt bagage Work 1, expression 1, manifestation 1

\*041 \$a swe  
 \*080 \$a 839.79-3  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Resa med lätt bagage \$b noveller \$d Tove Jansson  
 \*260 \$a [Hfors] \$b Schildt \$c 1987 \$e pain. Ruotsissa  
 \*300 \$a 173, [1] s. \$c 21 cm

Work1, expression 2, manifestation 1 Relationship: derivative : ny utg.  
 (= new edition)

\*041 \$a swe  
 \*080 \$a 839.79-3  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Resa med lätt bagage \$b noveller \$d Tove Jansson  
 \*250 \$a [Ny utg.]  
 \*260 \$a [Stockholm] \$b Bonnier \$c 1990  
 \*300 \$a 173, [1] s. \$c 18 cm  
 \*440 \$a Bonnier pocket

Work 1, expression 3, manifestation 1 Recording of reading of a literary work. Relationship: derivative:

\*041 \$a swe  
 \*080 \$a 839.79-3  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Resa med lätt bagage \$z Aänite \$d Tove Jansson  
 \*260 \$a Nykarleby \$b Studio FsR \$c 1996  
 \*300 \$z 5 C-kas. \$a (5 h 15 min)

Work 1, expression 4, manifestation 1. Relationship: derivative

\*0411 \$a ita \$c swe  
 \*080 \$a 839.79-3  
 \*1001 \$a Jansson \$h Tove  
 \*241 \$a Resa med lätt bagage  
 \*2452 \$a Viaggio con bagaglio leggero \$d Tove Jansson  
     \$e introduzione di Carmen Giorgetti Cima  
     \$e [traduzione dallo svedese di Carmen Giorgetti Cima]  
 \*260 \$a Milano \$b Iperborea \$c 1994  
 \*490 \$a Iperborea \$v 44 \$y 0044  
 \*70021\$a Giorgetti Cima \$h Carmen

Work 1, expression 5, manifestation 1. Relationship: derivative. Relationship: whole-part. -independent part

\*0411 \$a fin \$c swe  
\*080 \$a 839.79-3  
\*1001 \$a Jansson \$h Tove  
\*241 \$a Stenåkern  
\*241 \$a Resa med lätt bagage  
\*2452 \$a Kevyt kantamus ja muita kertomuksia \$d Tove Jansson  
\$e suom. Oili Suominen  
\*260 \$a Porvoo \$a Hki \$a Juva \$b WSOY \$c 1989 \$e Juva

Work: Jansson, Tove. Kometjakten. 1946 Published also under the name Kometen kommer, 1968

\*1001 \$a Jansson \$h Tove  
\*2452 \$a Kometjakten  
\*260 \$a Hfors \$b Schildt \$c 1946

Relationship: derivative.

\*1001 \$a Jansson \$h Tove  
\*2452 \$a Kometen kommer  
\*260 \$a Hfors \$b Schildt \$c 1968  
\*300 \$a 149 (2) s. \$b kuv.  
\*501 \$a Ilm. v. 1946 nimellä Kometjakten

Danish translation (expression) of the work Kometjakten. Relationship: derivative

\*1001 \$a Jansson \$h Tove  
\*241 \$a Kometjakten  
\*2452 \$a Kometen kommer \$d fortalt og ill. af Tove Jansson  
\$e overs. af Marie Svendsen  
\*260 \$a [København] \$b Carlsen \$c 1974

2 free translations into Danish of the work Kometjakten According to FRBR free translations are treated as new works Relationship: expression to expression between expressions of different works.

\*0411 \$a dan \$c swe  
\*080 \$a 839.79-3(024.7)  
\*1001 \$a Jansson \$h Tove  
\*2452 \$a Mumitrolden opdager kometen  
\$d tekst og tegninger: Tove Jansson  
\$e oversat af Børge Bavngård efter "Kometjakten"  
\*260 \$a København \$b Eiler Wangels forlag \$c 1954  
\*70011\$a Bavngård \$h Børge

\*0411 \$a dan \$c swe  
\*080 \$a 839.79-3(024.7)  
\*1001 \$a Jansson \$h Tove  
\*2452 \$a Mumitrolden vender hjem \$d tekst og tegninger af Tove Jansson  
\$e oversat af Børge Bavngård efter "Kometjakten"  
\*260 \$a København \$b Eiler Wangels forlag \$c 1954  
\*70011\$a Bavngård \$h Børge

Expression of the work in English embodied in two manifestations. Relationship: derivative. Relationship: Reprint statement indicates equivalent relationship.

#### Manifestation 1

\*1001 \$a Jansson \$h Tove  
\*241 \$a Kometjakten  
\*2452 \$a Comet in Moominland \$d written and ill. by Tove Jansson  
\$e transl. by Elizabeth Portch  
\*250 \$a Repr.  
\*260 \$a Harmondsworth \$b Penguin Books \$c 1973  
\*300 \$a 157 [1] s. \$b kuv. \$c 19 cm  
\*490 \$a Puffin Books  
\*500 \$a First publ. 1967  
\*555 \$a Repr. 1974. - Repr. 1979  
\*70011\$a Portch \$h Elizabeth

#### Manifestation 2

\*1001 \$a Jansson \$h Tove  
\*241 \$a Kometjakten  
\*2452 \$a Comet in Moominland  
\$d Tove Jansson \$e translated by Elizabeth Portch  
\*260 \$a [New York, NY] \$b Farrar, Straus and Giroux \$c 1991  
\*300 \$a 192 s. \$b kuv. \$c 19 cm  
\*490 \$a A Sunburst book  
\*70011\$a Portch \$h Elizabeth

Finnish expression of the work Kometjakten embodied in a multipart manifestation. Relationship: whole-part

-collection without collective title:

\*1001 \$a Jansson \$h Tove  
\*241 \$a Kometjakten  
\*241 \$a Trollkarlens hatt  
\*2452 \$a Muumipeikko ja pyrstötähti \$a Taikurin hattu  
\$d Tove Jansson  
\$e suomentanut Laila Järvinen \$e [tekijän kuvittama]  
\*745 \$a Taikurin hattu

-collection with collective title:

\*1001 \$a Jansson \$h Tove  
\*2452 \$a Takaisin Muumilaaksoon \$d Tove Jansson  
\$e kuv. Tove Jansson  
\*500 \$a Alkuteokset: Kometjakten ; Trollkarlens hatt ;  
Muminpappans bravader ; Farlig midsommar ;  
Det osynliga barnet och andra berättelser  
\*505 \$a Sisältö: Muumipeikko ja pyrstötähti ; Taikurin hattu ;  
Muumipapan villi nuoruus ; Vaarallinen juhannus ;  
Näkymätön lapsi / suom. Laila Järvinen

Work: Jansson, Tove. Det osynliga barnet och andra berättelser, 1962  
-some examples

Work 1, expression 1, manifestation 1

\*1001 \$a Jansson \$h Tove  
\*245 4\$aDet osynliga barnet och andra berättelser  
\$d Tove Jansson  
\*260 \$a Helsingors \$b Schildt \$c 1962

Work 1, expression 1, manifestation 2.

\*1001 \$a Jansson \$h Tove  
\*245 4\$aDet osynliga barnet och andra berättelser  
\$d Tove Jansson  
\*260 \$a Stockholm \$b Geber \$c 1962

Work 1, expression 2 (in Finnish), manifestation 1.

\*1001 \$a Jansson \$h Tove  
\*241 4\$a Det osynliga barnet och andra berättelser  
\*245 2\$aNäkymätön lapsi ja muita kertomuksia \$d Tove Jansson  
\*260 \$a Porvoo \$b [s.n.] \$c 1962

Work 1, expression 3 (in Finnish,) appearing in a multipart manifestation 1 with collective title. Relationship: whole-part

\*1001 \$a Jansson \$h Tove  
\*2452 \$a Takaisin Muumilaaksoon \$d Tove Jansson  
\$e kuv. Tove Jansson  
\*260 \$a Porvoo \$a Hki \$a Juva \$b WSOY \$c 1988 \$e Porvoo  
\*300 \$a 250, [1] s. \$b kuv. \$c 30 cm  
\*500 \$a Alkuteokset: Kometjakten ; Trollkarlens hatt ;  
Muminpappans bravader ; Farlig midsommar ;  
Det osynliga barnet och andra berättelser  
\*505 \$a Sisältö: Muumipeikko ja pyrstötähti ; Taikurin hattu ;  
Muumipapan villi nuoruus ; Vaarallinen juhannus ;  
Näkymätön lapsi / suom. Laila Järvinen

Work 1, expression 4 (in Finnish), manifestation 1. Recording of reading of the literal text:

\*1001 \$a Jansson \$h Tove  
\*2452 \$a Näkymätön lapsi \$z Äänite \$d Tove Jansson  
\*260 \$a [Helsinki] \$b WSOY \$c p1993  
\*300 \$z 6 C-kas. \$b Dolby

Work 2, expression 1 (in Finnish), manifestation 1. Recording of reading is based on the Finnish expression of the work. Relation: new work. Relation: whole-part

\*1001 \$a Jansson \$h Tove  
\*2452 \$a Muumilaaksossa tapahtui \$z Äänite  
\$d Tove Jansson \$e suomennos: Laila Järvinen  
\*300 \$z 1 C-kas. \$b Dolby  
\*440 \$a Ennenkuulumattomat  
\*500 \$a Teoksesta Näkymätön lapsi.

Work 3, expression:1 (in Finnish), manifestation 1. Recording of reading is based on the Finnish expression of the work Relation: new work. Relation: component part

\*021 \$a 951-0-92792-9  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Hattivattien salaisuus \$z Aänite \$d Tove Jansson  
     \$e suomentanut Laila Järvinen  
 \*300 \$z 1 C-kas. \$b Dolby  
 \*440 \$a Ennenkuulumattomat  
 \*500 \$a Teoksesta Näkymätön lapsi

Work 3, expression 1 (in Finnish), manifestation 2. Recording of reading is based on the Finnish expression of the work: Relation: whole-part: component part in a multipart manifestation with collective title .

\*021 \$a 951-0-25026-0 \$c koko teos  
 \*021 \$a 951-0-92701-5 \$c Hemuli joka rakasti hiljaisuutta  
 \*021 \$a 951-0-92381-8 \$c Kuka lohduttaisi Nytytiä?  
 \*021 \$a 951-0-92792-9 \$c Hattivattien salaisuus  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Hemuli, Nytyti ja hattivatit \$z Aänite  
     \$d Tove Jansson  
     \$e suomennos: Laila Järvinen ja Kirsi Kunnas  
 \*260 \$a [Helsinki] \$b WSOY \$c 2000\$f Kasettilinja  
 \*300 \$z 3 C-kas. \$b Dolby HX PRO  
 \*440 \$a Kasettikirjat  
 \*745 \$a Hemuli joka rakasti hiljaisuutta  
 \*745 \$a Kuka lohduttaisi Nytytiä?  
 \*745 \$a Hattivattien salaisuus

Work 4, expression 1, (in Norwegian), manifestation 1. Relation: adaptation, based on the TV-serie of Moomins. Relation: new work. Relation: component part in a multipart manifestation without collective title.

\*100 \$a Jansson, Tove  
 \*2401 \$a Alicia \$a Ninni  
 \*24500\$a Trollpiken Alicia ; Det usynlige barnet  
     \$c av Tove Jansson ; oversatt og bearbeidet av Henning Hagerup  
 \*260 \$a [Oslo] \$b Cappelen \$c 1997  
 \*500 \$a Oversatt etter svensk utgave som bygger på  
     den japanske TV-serien om Mummitrollet  
 \*7400 \$a Mummitrollet møter trollpiken Alicia  
 \*7400 \$a Tove Janssons Mummitrollet møter trollpiken Alicia  
 \*7400 \$a Trollpiken Alicia  
 \*7404 \$a Det usynlige barnet  
 \*991 \$a Hagerup, Henning

## F Example records: Tove Jansson

Here are some examples causing incorrect statistics:

- different cataloguing traditions
- misspellings and mistakes
- incorrect cataloguing

- there is no duplicate control for material extracted from different databases.

## Examples

Muumintrollet (10 works/expressions embodied in 36 records) by Tove Jansson exists as three different "works" due to the different cataloguing traditions.

Mumintrollet (in Finnish: Muumipeikko) is the common title for a multivolume publication consisting of 10 parts. Each individual part of the work Mumintrollet contains separate works/expressions. So the work/expression as a part of multi-volume publication is at the same time a manifestation of several separate works. Some parts might also be published as new editions.

As a result of machine extraction the work "Mumintrollet" exists as 4 different works on the list of works.

```
Mumintrollet
Mumintrollet 1
Innehålder episoderna
Innehålder episoderna : muminvinter

*1001 $a Jansson $h Tove
*2452 $a Mumintrollet $d av Tove Jansson
*248 $g 4
*260 $a [Stockholm] $b Norstedt
*260 1$c 1989 $e Keuruu $f Otava
*505 1$a Innehåller episoderna: Mumintrollet blir förälskat ;
      Mumintrollet och marsinnevånarna ;
      Muminfamiljen och havet
*745 $a Mumintrollet blir förälskat
*745 $a Mumintrollet och marsinnevånarna
*745 $a Muminfamiljen och havet
```

The titles in the content note field are not extracted mechanically.)

```
*1001 $a Jansson $h Tove
*2411 $a Mumintrollet 4
*2452 $a Muumipeikko $d Tove Jansson
*248 $g 4 $d [suomennos: Juhani Tolvanen ja Anita Salmivuori]
      $y 0004
*260 $a Porvoo $a Helsinki $a Juva $b WSOY
*260 1$c 1990 $e Juva
*505 1$a Sisältää jaksot: Muumipeikko rakastuu ;
      Muumipeikko ja marsilaiset ; Muumiperhe ja meri
*555 $a 2.-3. p. 1992
*70011$a Tolvanen $h Juhani
*70011$a Salmivuori $h Anita

*1001 $a Jansson $h Tove
*2452 $a Mumintrollet $d Tove Jansson
*248 $g 6 $h Innehåller episoderna
      $b Föreningsliv och Den gyllene svansen
```

```

*260 $a Stockholm $b Geber
*260 1$c 1962

*1001 $a Jansson $h Tove
*2452 $a Mumintrölet $d Tove Jansson
*248 $g 9 $h Innehåller episoderna: Knappar och äktenskap,
      Dam i dilemma
      $d av Lars Jansson
*260 $a Stockholm $b Geber $a Helsingfors $b Schildt
*260 1$c 1965
*300 1$a 92 s. $b kuv.
*70011$a Jansson $h Lars

```

In addition 41 expressions of the work are listed erroneously as separate works because of -omission of the original title information -or because of human mistakes in cataloguing, e.g. :

- the lack of original title information
- the original title information is given in notes
- the original title information is incorrect
- different works with identical names

### Examples

-original title of the work is missing (24):

```

*001f563742
*015 $a f563742
*0411 $a eng $c swe
*080 $a 839.79-3(024.7)
*1001 $a Jansson $h Tove
*2452 $a Who will comfort Toffle? $d Tove Jansson
      $e translated by Kingsley Hart
*260 $a London $b Ernest Benn $c 1960
*300 $a [30] s. $b kuv. $c 4:o
*70011$a Hart $h Kingsley

```

-original title information is "hidden" in description (4):

```

*015 $a f577900
*0411 $a dan $c swe
*080 $a 839.79-3(024.7)
*1001 $a Jansson $h Tove
*2452 $a Munitrollden vender hjem $d tekst og tegninger af Tove Jansson
      $e oversat af Børge Bavngård efter "Kometjakten"
*260 $a København $b Eiler Wangels forlag $c 1954
*300 $a 69 s. $c 8:o
*70011$a Bavngård $h Børge

```

-incorrect original title information. The original title should be: Det osynliga barnet och andra berättelser(2):

\*021 \$a 82-03-11199-8 \$c sid.  
 \*0411 \$a nor \$c swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*241 \$a Muminpappans bravader  
 \*24524\$a Det usynlige barnet og andre fortellinger  
       \$d Tove Jansson \$e overs. av Gunnel Malmström  
 \*250 \$a 3. oppl  
 \*260 \$a Oslo \$b Aschehoug \$c 1983  
 \*300 \$a 132, [1] s. \$b kuv. \$c 22 cm  
 \*70011\$a Malmström \$h Gunnel

-different works with identical names interpreted as the same work (3):

\*021 1\$a 91-7556-011-9 \$c nid. \$d 21 mk  
 \*041 \$a swe  
 \*080 \$a 741.5  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Mumin \$d av Tove Jansson och Lars Jansson  
 \*248 \$g 6

\*021 \$a 952-9549-05-9 \$c nid.  
 \*0411 \$a ger \$c swe  
 \*080 \$a 7.091  
 \*080 \$a 73/76  
 \*1001 \$a Jansson \$h Tove  
 \*2452 \$a Mumin \$b das Mumintal, Kunstmuseum Tampere  
       \$d [Texte, Zeichnungen und Aquarelle: Tove Jansson]  
       \$e [photographische Arbeiten: Per Olof Jansson]

\*021 \$a 91-7100-181-6 \$c nid.  
 \*041 \$a swe  
 \*080 \$a 7.091  
 \*080 \$a 74  
 \*1001 \$a Jansson \$h Tove  
 \*245 2\$a Mumin \$b Nationalmuseum 24 oktober 1980 -  
       6 januari 1981 \$d illustrationer [och text:]  
       Tove Jansson \$e foto: Per Olof Jansson

-misspellings causing "new works" (5):

\*015 \$a f563723  
 \*0411 \$a nor \$c swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*241 \$a Farlig midsommer  
 \*2452 \$a Farlig midtsommer \$d Tove Jansson  
       \$e oversatt etter ... av Gunnel Knudsen Malmström

\*015 \$a f568223  
 \*0411 \$a eng \$c swe  
 \*080 \$a 839.79-3(024.7)  
 \*1001 \$a Jansson \$h Tove  
 \*241 \$a Farling midsommar  
 \*2452 \$a Moominsummer madness \$d written and illustrated



by Tove Jansson  
\$e translated by Thomas Warburton

- or language codes have switched places (3):

\*001f568220  
\*008930511s1957 dk jp f  
\*015 \$a f568220  
\*035 \$a 0368-57260  
\*0411 \$a swe \$c dan  
\*080 \$a 839.79-3(024.7)  
\*1001 \$a Jansson \$h Tove  
\*241 \$a Hur gick det sen?  
\*2452 \$a Hvad sker der så? \$b [bogen om Mumi, Mymle og lille My]  
\$d Tove Jansson \$e [på dansk ved Ole Sarvig]  
\*260 \$a København \$b Gyldendal  
\$c [1957] \*300 \$a [26] s . \$b kuv. \$c 4:o  
\*70011\$a Sarvig \$h Ole

## References

- [1] Trond Aalberg and Knut Hegna. *Arkitektur for digitale bibliotek*. BIBSYS, 2000. Architecture for digital libraries. [In Norwegian]  
<http://www.bibsys.no/BDB/arkitektur/ArkDigBib.html>.
- [2] James Duff Brown. *Library classification and cataloguing*. Grafton & Co., 1916.
- [3] Charles A. Cutter. *Rules for a dictionary catalog : special report on public libraries*. Government printing office, 1904.
- [4] Tom Delsey. Functional analysis of the MARC 21 Bibliographic and Holdings Formats. Technical report, Library of Congress, 2001. Prepared for the Network Development and MARC Standards Office, Library of Congress  
<http://www.loc.gov/marc/marc-functional-analysis/home.html>.
- [5] Bernhard Eversberg. Reuse+ - the Part à Whole Relationship in German and American Cataloging Data : Results and suggestions. Appendix: Links between bibliographic records. Technical report, Joint Project of OCLC and Niedersächsische Staats- und Universitätsbibliothek Göttingen, 27 June 1998.  
<http://www.biblio.tu-bs.de/allegro/formate/reusep.htm>.
- [6] Format Variation Working Group. Interim report. Technical report, Joint Steering Committee for the revision of the Anglo-American Cataloging Rules, 2001/10/09.  
<http://www.nlc-bnc.ca/jsc/forvarwg3.pdf> via  
<http://www.nlc-bnc.ca/jsc/current.html>.
- [7] Knut Hegna. MARC study tool : user guide. Technical report, University of Oslo, Informatics Library, 2002.  
<http://folk.uio.no/knuthe/dok/userguide.pdf>.
- [8] IFLA Study group on Functional requirements for bibliographic records. Final report. Technical report, IFLA, 1998.  
<http://www.ifla.org/VII/s13/frbr/frbr.pdf>.
- [9] Linnea Database Services. *FINMARC bibliographic format : version 14.8.2000*. Helsinki University Library. The National Library of Finland, 2000. The English version is available via  
<http://www.lib.helsinki.fi/kirjastoala/linnea/formaatti.htm>.
- [10] The Norwegian Cataloguing Committee, Oslo. *NORMARC*, 1998.  
via <http://www.nb.no/katkom/>.
- [11] Elaine Svenonius. *The intellectual foundation of information organization*. MIT press, 2000.
- [12] Barbara B. Tillett. Bibliographic relationships. In Carol A. Bean and Rebecca Green, editors, *Relationships in the organization of knowledge*, chapter 2, pages 19 – 35. Kluwer academic publishers, 2001.