UNIVERSITY OF LJUBLJANA
FACULTY OF ARTS
DEPARTAMENT OF LIBRARY AND INFORMATION SCIENCE
AND BOOK STUDIES

Mirlona Buzo

THE CONVERSION OF THE BIBLIOGRAPHIC RECORDS
FROM THE WINISIS FORMAT
INTO THE COMARC FORMAT

Master's thesis

with appendices on CD-ROM

Ljubljana, 2012

UNIVERSITY OF LJUBLJANA
FACULTY OF ARTS
DEPARTAMENT OF LIBRARY AND INFORMATION SCIENCE
AND BOOK STUDIES

Mirlona Buzo

THE CONVERSION OF THE BIBLIOGRAPHIC RECORDS
FROM THE WINISIS FORMAT
INTO THE COMARC FORMAT

Master's thesis

with appendices on CD-ROM

Supervisor: assoc. prof. dr. Alenka Šauperl   Study program: Library Science
Co-supervisor: assist. prof. dr. Marija Petek

Ljubljana, 2012

Abstract

The purpose of this thesis is to prepare a model for converting bibliographic data from the library local format into the COMARC format. The model will be developed on the basis of the data from the catalogue of the Academy of Sciences of Tirana. With the use of various structural and functional methods, records will be analysed and tables prepared for mapping the local format into the COMARC format. After the mapping, the analysis of the record quality will follow. For this purpose, the actual records will be compared and contrasted with the ideal record in the test environment. The ideal record will be created in accordance with the requirements of the ISBD standard, the COMARC format and the second level of bibliographic description according to the AACR2 cataloguing rules. The results of this analysis will be used to improve the model for upgrading the records until the required quality is reached and is in accordance with the above mentioned standards and recommendations. Another purpose of this thesis is to prepare a universal model that could be used in the future for converting the catalogues of other Albanian libraries provided with the CDS/ISIS systems.

Translation into Slovenian:

Naslov: KONVERZIJA BIBLIOGRAFSKIH ZAPISOV IZ FORMATA WINISIS V FORMAT COMARC

Ključne besede: bibliografski zapisi, konverzija, formati za preslikavo, format COMARC, format WINISIS

Izvleček

Namen te naloge je pripraviti model za pretvorbo bibliografskih podatkov iz lokalnega formata knjižnice v format COMARC. Model bo razvit na podlagi podatkov iz kataloga Akademije znanosti v Tirani. Z uporabo različnih strukturnih in funkcionalnih metod bodo zapisi analizirani. Pripravljene bodo tabele za preslikavo iz lokalnega formata v format COMARC. Po preslikavi bo sledila analiza kakovosti zapisov. V ta namen se bo v testnem okolju uporabila metoda primerjanja idealnega zapisa z dejanskimi zapisi. Idealen zapis bo izdelan v skladu z zahtevami standarda ISBD, formata COMARC in katalogizacijskimi pravili AACR2 za srednji nivo bibliografskega opisa. Rezultati te analize se bodo uporabili za izboljšanje modela za nadgradnjo zapisov, dokler ne bo dosežena raven kakovosti, ki jo zahtevajo zgoraj navedeni standardi in priporočila. Prav tako je namen te naloge pripraviti splošen model za konverzijo, ki bi se lahko v prihodnosti uporabil za pretvorbo katalogov drugih albanskih knjižnic, ki uporabljajo sisteme CDS/ISIS.

Acknowledgement

I would like to express my very sincere gratitude to my supervisor **assoc. prof. dr. Alenka Šauperl** and co-supervisor: **assist. prof. dr. Marija Petek**, who offered their continuous advices and encouragements throughout the guidance of this thesis.

I would like to express my sincere thanks to my colleague Darja Boksa-Faraguna for editing/translation of my thesis.

I am thankful to my colleagues of IZUM for the support during my stay in Maribor and for their technical assistance to my thesis.

Last but not the least; I would like to thank my family who supports me spiritually in my life.

CONTENTS

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CDS/ISIS - Computerized Documentation System - Integrated Set for Information Systems

WINISIS - CDS/ISIS for Windows

MARC - Machine Readable Cataloguing Records

UNIMARC - Universal MARC format

COMARC - Cooperative Machine-Readable Cataloguing

COMARC/B - Cooperative Machine-Readable Cataloguing for bibliographic description

COMARC/H - Cooperative Machine-Readable Cataloguing for Holdings Data

COBISS - Co-operative Online Bibliographic System & Services

ASAL - The Academic of Science of Albania Library

UNESCO - the United Nations Educational, Scientific and Cultural Organization

UDC - Universal Decimal Classification

ISBD - International Standard Bibliographic Description

AACR2 - Anglo-American Cataloguing Rules, Second Edition

LCSH - Library of Congress Subject Headings

OCLC - Online Computer Library Centre

1 INTRODUCTION

Libraries and information centres are the intermediaries between information, information sources and users. In order to make information accessible, libraries perform several activities. Among them, cataloguing is one of the most important and fundamental activity to make information accessible (Chudamani & Nagarathna, 2006). In the era of global resource sharing, the conversion of bibliographic records from one system to another is imperative in library communities.

During conversion one form of data is transferred to another form. This is called data migration which itself is considered to be the most problematic aspect of the changeover of a library management system (Hallmark & Garcia, 1992), especially for libraries switching from a first generation system to a new system.

Nowadays, the Universal MARC format (UNIMARC) is most often used for the processing of bibliographic material. In the recent years, it has been necessary to convert data to UNIMARC to meet the needs for exchange of bibliographic data within the UNIMARC-based global library network. In practice, a variety of these formats are used (Rudic & Surla, 2008). Hence, a version of UNIMARC, called COMARC (Cooperative Machine Readable Cataloguing), is used in the COBISS (Co-operative Online Bibliographic System & Services) library software.

Thousands of small to medium-sized libraries in developing countries continue to use CDS/ISIS (Computerised Information Service/Integrated Scientific Information System) databases for their daily operations. CDS/ISIS is a family of non-relational database systems with a history that goes back to the 1970's. It was developed specifically for bibliographic data by UNESCO (United Nations Educational, Scientific and Cultural Organisation), and is still officially distributed and widely used (Hopkinson, 2005). The CDS/ISIS programme has positively impacted libraries with limited resources. It has allowed them to computerise their catalogues using simple stand-alone PCs. Some Web-enabled CDS/ISIS products have become more and more common; however, progress has been stalling over the last ten years.

Data migration is problematic in the sense that libraries use different standards of cataloguing. Instead of using a single standard to create machine-readable catalogue records, a lot of standards have emerged and are today used by different institutions. Because of these variations in standards, sharing

of resources and transfer of data from one system to another among the institutions, locally and globally, has become a significant problem (Chandrakar, 2002).

The Library of the Academy of Sciences of Albania (ASAL) has implemented CDS/ISIS (Version 1.3) software for Windows in a local MARC format version provided by UNESCO since 2004. In 2010, this software was updated to the new version (WINISIS 1.5, build 3). The library has created a total of 6838 bibliographic records for newly entered books. Additionally, cataloguing rules (in general) have gone through many changes. ASAL uses the Universal Decimal Classification Scheme (UDC), including Extensions and Corrections to the UDC. For cataloguing purposes, they also use International Standard Bibliographic Description for monographic publications, ISBD (M), and Anglo-American Cataloguing Rules, Second Edition (AACR2). Subject cataloguing is a relatively new experience for Albanian librarians. It started back in 1995 based on the list of subject headings, which has been the translation and adoption in the Albanian language of the Sears List of Subject Headings. Realizing that the List was restricted for the collection of the National Library, cataloguers decided to use Library of Congress Subject Headings (LCSH); they made the necessary translations into Albanian and tailored headings to suit their needs.

This thesis describes the comparison between bibliographic catalogue records of ASAL and bibliographic records of COMARC/B. It also describes the approach used for tackling problems that occur during the mapping or exchanging of records, as well as data migration from WINISIS to COMARC.

1.1   OBJECTIVES

1.1.1   Problem description

The rapid growth of technology has forced libraries to adapt by striving to keep up with technological advances. Catalogues need to respond to the demands of users. This is why the libraries should either update their bibliographic record format or migrate their databases to other library system.

There are a number of reasons why any library would wish to switch over from one library management system to another (Matoria & Upadhyay, 2004, p. 161).

Some of them are:

- Phasing out the existing software by vendor

- Lack of required features in existing version

- Cost of maintenance of the existing software

- Change forced due to the technological advancement

- Integration of the database to an external system that uses another software

When a library attempts a mapping between two standards, the basic question that needs to be answered is "Why mapping? (Chandrakar, 2005, p. 588) ". The reasons for the mapping are:

- Universal availability of resources – Organisations use standards to create bibliographic machine-readable records, but the basic structures of the formats used are different, which does not provide uniformity within records. In order to avoid this problem, mapping is required to serve the purpose of universal availability of resources.

- The development of metadata makes libraries to update their format for bibliographic description in order to be coherent and to support interoperable data exchange.

- Searching across syntaxes and databases – Libraries have large systems with valuable information in metadata records. With the expansion of electronic resources on the Internet, it will be necessary for such systems to be able to search metadata across different syntaxes and databases, and to identify commonality in the definition and use of elements.

In 2004, ASAL began building its electronic catalogue based on the WINISIS program; since then, however, it has not updated or taken on any maintenance work in terms of the bibliographic record format. Meanwhile, libraries worldwide have managed to develop and improve their bibliographic record formats from MARC to UNIMARC or MARC 21. It is for this reason that it has become necessary for the ALAS library to start upgrading its catalogue format.

On one hand, the WINISIS format contains a limited number of fields and subfields. The conversion could provide the expansion of fields and subfields structure and thus enable more detailed information to be added.

On the other hand, the WINISIS program used by the ASLA library is a local program. The users do not have online access to library materials outside the library's walls, which makes the library isolated from other library systems. The conversion of the database to the COMARC (COBISS programme) format would provide not only online access for the users, but also shared cataloguing membership and the availability of the information at the regional level and beyond.

Librarians (cataloguers), by taking part in shared cataloguing, become aware of the ideal quality of cataloguing and they may, as well, be able to compare their bibliographic records with others, or edit them on the basis of the international standards.

The conversion is done automatically if 1) the WINISIS records meet at least minimum quality standards, 2) the process does not require too much manual record "cleaning" and updating, and 3) there is no possibility to acquire higher quality records from other databases. Of course, a conversion process has its cost. Automatic conversion is cheaper and less time-consuming if compared with manual data entering for a second time (Phillips, 1999).

Even if many librarians would prefer re-cataloguing, i.e. creating "perfect" records in full conformity with the existing cataloguing rules, in most conversion projects pragmatic reasons, such as time and cost, restrict this approach to a minority of "problematic" records (Phillips, 1999). The library staff needs a lot of time to re-catalogue such an amount of bibliographic records manually.

These processes are time and labour consuming. Also, such conversions are inefficient and cause errors to come up again and again. In automatic conversions, however, errors occur due to incorrect code or wrong algorithm. The errors can then be corrected once they are identified, which means that they are not recurring in nature.

When data are converted manually, errors recur as they happen, at different points of time, with many people involved in the process and a lot of human intervention. In the process, automatic conversions minimise human interventions and with it decrease the probability of errors through time, hence, they do not only save time and labour but also minimise errors.

1.1.2   Scope

This research provides a model for converting bibliographic data from the WINISIS local library format to COMARC/B format. The model will be developed on the basis of the data from the ASAL catalogue in order to build a universal model for conversion, which, in the future, may be used for converting the catalogues of other Albanian libraries that use the same CDS/ISIS system.

1.1.3   Research questions

In order to achieve the main objective, the research questions are set out below:

1. How compatible are the two formats?

First, by analysing and discussing the similarities and differences of the catalogue record structure between the two systems, we will have to determine which WINISIS/B fields and subfields correspond to COMARC/B fields and subfields.

The mapping of data will give us an idea about the amount of work necessary for the conversion. It will tell us whether the conversion can be automated fully or partially. Consequently, the fields that will not be directly mapped will be added, changed or deleted. The numbers of such fields will show how different the two formats are. It is expected that the two formats are similar in the ISBD standard, and partially similar in the structure of the formats.

2. How much will bibliographic records improve?

We will try to determine the improvement of the records by comparing the actual records with the ideal record in accordance to international standards. The comparative analysis will help us determine how to correct the records during the conversion process. From the analysis of the local database records we expect to improve data during the conversion process by adding the fields/subfields that are not present in the WINISIS format and by correcting the data through the conversion process.

3. What quality will the converted bibliographic records be?

The converted records will be verified automatically to detect possible errors. The converted bibliographic data will be compared with the records from well-established databases, such as WorldCat, OCLC or COBISS.Net. The comparison will highlight the necessity to improve the records in order to

achieve the required quality level. From this comparison we expect that the converted records will have almost the approximate quality of the other standard's records. The comparison will emphasise the necessity to improve the records at the required level that meets the ISBD/M standard, the second level of the bibliographic description of the AACR2 1.0D2 rules and the COMARC/B format.

4. How efficient is the conversion from WINISIS to COMARC/B?

Finally, the study must come to a conclusion about the efficiency of the conversion and the generality of the method, the proof that the provided method is either applicable to other conversions, or it is just an ad-hoc solution efficient only for the given formats. It is expected to have an automatic conversion process and we attempt not to lose data during this process.

1.1.4   Study significance

This study presents an approach that enumerates problems that may occur during the mapping and exchanging of records. It will also tell us how to overcome these problems. Through a real-world success story it makes libraries commit to considering the COBISS solution, along with several other factors. The libraries that will consider moving from WINISIS to COBISS will, in particular, stand to benefit from this experience.

If the study shows that the quality of converted records is feasible and sufficient, the records will be used to build the initial COBIB.AL shared database and the local databases of the Academy of Sciences of Tirana within the COBISS.AL system.

This study will have implications for libraries still operating under legacy systems. It will be expected from ASAL, in particular, to acquaint all these libraries in Albania with its own experience and, hopefully, they will, sooner or later, migrate to a new system as well. Despite all difficulties, libraries in Albania will be encouraged to follow the ASAL example, and in doing this, plan to acquire their own library system, and eventually decide for data migration.

In the future, this study will prove useful in importing/exporting data from software, such as CDS/ISIS that is used by Albanian libraries. Moreover, the switch over to software will be beneficial for libraries in terms of being able to refine and clean their existing data during the conversion process.

1.1.5 Definition of terms

Some of the terms used in this thesis refer to librarianship and information science (field, subfield, mapping table), whereas others are strictly technical terms (conversion algorithm, ideal record).

Field. Each bibliographic record is logically divided into fields. There is a field for the author, a field for title information, and so on. These fields are subdivided into one or more **subfields**. As previously noted, the textual names of the fields are too lengthy to be reproduced within each MARC record. Instead, they are represented by 3-digit codes.

MARC record. This is a Machine-Readable Cataloguing record. **Cataloguing record** means a bibliographic record, or the information traditionally shown on a catalogue card. The record includes (not necessarily in this order): 1) a description of the item, 2) main entry and added entries, 3) subject headings, and 4) the classification or a call number. MARC records often contain additional information.

Mapping. A correspondence specification between two metadata standards. The development of mapping is the intellectual task of determining the semantic mapping of elements between the source and target metadata standards. The task involves specifying the mapping of each element in the source metadata standard with a semantically equivalent element in the target metadata standard. The prerequisite to a meaningful mapping requires a clear and precise definition of the elements in each standard.

Conversion algorithm. An algorithm is a short and strict definition to tell a programmer what to do with each field and subfield in the conversion process. It is used when something wrong is repeated.

Ideal record. A record created in accordance with the requirements of the ISBD/M standard, the COMARC/B format and the second level of bibliographic description according to Anglo-American Cataloguing Rules (ACR2), Rule 1.0D2.

Record quality. A record that has enough description to identify an item, enough access points to enable record identification and retrieval, as well as the appropriate detail to users for the relevant function to be performed.

1.2 DESCRIPTION OF STUDIED FORMATS

1.2.1 Characteristics of the COMARC/B format

The COMARC/B format for bibliographic description and data exchange within the Co-operative Online Bibliographic Systems and Services (COBISS system) is based on the UNIMARC bibliographic format, which is the international standard developed until March 2003 by IFLA Universal Bibliographic Control and International MARC Programme (UBCIM) based in Deutsche Bibliothek, Frankfurt am Main. In March 2003, the UBCIM Programme was closed and the National Library of Portugal (NLP) took over the development of UNIMARC (COMARC/B format for bibliographic data, 2008).

Within the COBISS system, the COMARC/B format for bibliographic data includes, beside the UNIMARC standard fields and subfields, a number of specific features of the bibliographic record structure. COMARC/B contains 147 fields, mainly adopted from UNIMARC, and 21 COMARC-specific fields. Each COMARC record contains an automatically created "system field" that contains details about who created and updated the record, the dates of record manipulation, record version, etc.

In addition to COMARC/B, there are two other formats used within the COBISS system: COMARC/H for holdings data and COMARC/A for authorities data. The development of new functions within the COBISS system necessitated the upgrading of UNIMARC; new fields and subfields were added, which eventually led to the development of the COMARC/B format.

Fields are designated by a string containing three numerals, whereas subfields by a letter or a numeral. The smallest information unit being recognized by the program is a subfield (search, display), i.e. a subfield element in COMARC/H format for holdings data.

Some fields have the indicator values defined usually signifying a display mode. In some cases fields and subfields are repeatable.

To exchange data in the COBISS system, COMARC/B is used for bibliographic data, COMARC/A for authority data, and COMARC/H for holdings data. The first two formats are based on UNIMARC; the third was developed entirely by IZUM. For the international data exchange, the MARC 21 format is used.

Conversions are possible from COMARC into MARC 21 and vice versa. Records are available in either the ISO 2709 structure (MARC 21, COMARC) or XML (Dublin Core, MODS, MARC 21, COMARC). Bibliographic data can be exported from the COBISS system in MARC 21 and Dublin Core.

1.2.1.1 Other implementation of COMARC within COBISS

COBISS is a reference model of a system representing the platform for the national library information systems of Slovenia, Serbia, Macedonia, Bosnia and Herzegovina, Montenegro and Bulgaria which are interconnected in the COBISS.Net regional network. While Albania and Kosovo are in the process of preparing themselves for the implementation of the COBISS system, an open invitation for cooperation has been offered to libraries in Croatia (Györkös & Seljak 2006).

COBISS.SI, COBISS.SR, COBISS.MK, COBISS.BH, COBISS.CG and COBISS.BG are designations for the autonomous library information systems with their own COBIB shared bibliographic databases that have been, or are being, set up in individual countries (Slovenia, Serbia, Macedonia, Bosnia and Herzegovina, Montenegro and Bulgaria). As part of the aforementioned systems, over 650 libraries use COBISS software for the automation of their operations (COBISS Platform).

In 1996, IZUM first restored its co-operation with libraries in Macedonia, where currently the latest version of COBISS2 software is used by the four largest libraries, interconnected in the shared cataloguing system COBISS/MKD (Seljak & Seljak, 2002).

In 1998, the COBISS.BiH independent shared cataloguing system was established in Bosnia and Herzegovina. The project was financed in the context of the Slovenian Technical Support Donation Programme for Bosnia and Herzegovina. Following the model of IZUM for COBISS in Slovenia, the COBISS Centre in Sarajevo manages the central functions and services of the COBISS.BiH system. The COBISS Centre in BiH was founded by five universities in Bosnia and Herzegovina, and by IZUM (Seljak & Seljak, 2002).

In Serbia, IZUM restored co-operation with two libraries in 1997, and with another two in 2001. In co-operation with the Soros (Open Society Foundation) foundation in Belgrade and the organization Alternativna akademska obrazovna mreža (Alternative Academic Educational Network, Belgrade), the project "Virtual Library of Serbia" was completed in April 2001. This involved the interconnecting of libraries in Serbia into the COBISS independent shared cataloguing system.

A similar project has also been prepared for libraries in Montenegro, where currently the latest version of COBISS2 software is used by the two largest libraries, interconnected in the shared cataloguing system COBISS/CG.

In Croatia, the National and University library in Pula uses COBISS software (the 1991 version).

Reliable online communications between local systems and the central system is a precondition for the successful operating of the COBISS applications. This presents a huge obstacle to the implementation of shared cataloguing in these countries because of the underdevelopment of the communications infrastructure in the academic and research Internet networks (Seljak & Seljak, 2002).

1.2.2   Characteristics of the WINISIS format and its use in libraries

CDS/ISIS is an information retrieval package developed by UNESCO and distributed free of charge to non-profit organisations throughout the world. This factor is important, in particular, for developing countries where the cost implications are often critical. CDS/ISIS has enabled many developing countries to join and contribute to the information society (Keyser, 2000).

CDS/ISIS is a software package which has evolved from the original DOS version to a Windows version, and implements fields and subfields in the MARC format (Hopkinson, 1997).

One of the advantages of ISIS is that it accommodates the ISO-2709 format-based international information, such as MARC and CCF (Common Communication Format); (Curwen, 1997). These formats promote the exchange of bibliographic data. MARC is a standard which takes the bibliographic data elements found in traditional cataloguing rules (such as the Anglo-American Cataloguing Code) and defines a place for them in a record structure.

WINISIS differs in more than one way from other standards, e.g. it uses different subfield tags, different field and record terminators, and every record is divided into 80-character segments. Data elements may be stored in fields or subfields, each of which is assigned a numeric tag indicative of its contents.

A field, which may be optional (i.e. it may be absent in one or more records), may contain a single data element, or two or more variable length data elements. Furthermore, a field may be repeatable, i.e. any given record may contain more than one instance or occurrence of the field (WINISIS Manual, 2003).

One of the disadvantages of the WINISIS structure is a 2-character code subfield delimiter preceding and identifying a variable length subfield within a field. It consists of the character ^ (not sign) followed by an alphabetic or numeric character, e.g., ^a. The imprint field is defined as containing the information on the place of publication, publisher and date of publication in the three subfields a, b and c respectively. E.g. ^aParis^bUnesco^cl985. Many times librarians forget this character (^), which causes errors in the records.

WINISIS software is limited in what it can do, and it is not a total library automation system. It is not a completely suitable solution and this is seen as a disadvantage, as well. CDS/ISIS software is not user friendly.

WINISIS training was an important factor, which indicated that advanced training was required. Computer literacy alone was not sufficient for the use of WINISIS, however, UNESCO does not have the capacity to provide in-depth training.

1.2.2.1 Overview of WINISIS implementation in Albania

In 1989, the National Library installed CDS/ISIS software. The software was distributed free of charge by UNESCO as an information program to help information centres (including libraries) in Southeast Europe.

IT specialists in collaboration with specialists from the National Library adapted this program to build an electronic database, i.e. the local online catalogue.

At that time at the National Library, there were only a few computers, and were not connected in the local area network. The computers were provided with microprocessors DX version 3.06, version in DOS. At the Library, they began building a database in ISIS in order to catalogue books. At

the beginning, ISIS served as an experimental database used for the training of their librarians; later on, it was used also for the training of librarians from other libraries for the automation of information (Xhaja, 2001).

Automation efforts in other libraries in Albania started in 1995 through the project *Foundation Library Program Soros in Albania* (OSFA) in collaboration with the Albanian Libraries Association (ALA). From this project, 14 public libraries in the cities benefited: Shkodër, Korçë, Vlorë, Durrës, Fier, Elbasan, Berat, Lushnje, Sarandë, Peshkopi, Lezhe, Gjirokastër, Pogradec, Public Library no. 5 in Tirana, etc. All these libraries were provided with CDS/ISIS.

In 1998, the UNESCO program was upgraded to a new version of DOS switched to Windows and was called WINISIS. At this stage, the National Library reviewed once again the data structure, and added necessary fields according the ISO 2709 standard.

In the second phase, the Soros Foundation Library Programme expanded the distribution of computers to public libraries in Permet Lushnje, Puke, Krujë, Burrel, Librazhd, Rrëshen, Gramsh, as well as to several university libraries. Also this time, training workshops and programme installations followed at the National Library Training Centre.

WINISIS 1998 was installed also in about 40 libraries, among which were academic libraries, university and special libraries. Nevertheless, the libraries had problems creating their catalogues, mainly due to:

- Insufficient funding from the government;
- Frequent changing of personnel (insufficient wages and administrative reforms);
- Lack of IT staff at libraries responsible for the maintenance of equipment and program.

In these libraries, the automation of information started with new entries, retrospective conversion of data followed.

In 2004, the Academy of Sciences of Albania implemented CDS/ISIS Version 1.3 for Windows in the local MARC format provided by UNESCO.

1.2.2.2 The WINISIS data structure for this particular study

In the WINISIS format, fields are arranged in blocks. A field code consists of one, two or three characters. Only certain fields contain subfields.

WINISIS contains no indicators, repeatable fields and subfields are unique. The main characteristic of this format is no control over the database because it does not verify data entries. Records of monograph bibliographic description consist of fields and subfields. Each field has its own code (with one, two or three digits) (Keyser, 2000). The field's definitions and the bibliographic descriptions of the WINISIS database are presented in the *Table 1.1*. A detailed image of WINISIS data structure, data extraction, including data meaning, is prepared to provide maximum information of the library catalogue format (ASAL format/B).

The format used in this program is based on UNIMARC. However, ASAL has selected some fields and subfields from a list, provided by WINISIS, in the bibliographic description at a simple level, beginning from field 6 to field 922, and also a minimum of subfields. The library has also created some fields for their own administrative data, for example field 6 for "*call number*" and field 992 for "*type of document*".

The fields are sorted according to alphanumeric order, and include either repeatable or non-repeatable field type *(Table. 1.1)*.

Table 1.1 *Field definition table. Bibliographic descriptions of ASAL books 2004–2011*

Tag	Field type: R/NR	Field description	Subfield and description
6	NR	Call number	
10	R	International Standard Book Number	ˆa - ISBN ˆb - qualification ˆc - term of availability and/or price ˆz - wrong ISBN
101	NR	Language of the Item	^a - language of text ^c - language of original work
200	NR	Title and statement of responsibility	ˆa - title proper ˆe - other title information ˆh - number of part ˆi - name of part
205	NR	Edition statement	^a - edition statement
210	R	Publication	ˆa - place of publication ˆc - name of publisher ˆd - date of publication
215	NR	Physical description	ˆa -specific material designation ˆc - other physical details ˆd - dimension ˆe - accompanying material
225	NR	Series	ˆa - series title ˆx - ISSN of series ˆv - volume of designation
300	R	General note	
301	R	Shelf mark	
302	R	Location	
330	NR	Abstract	
600	R	Personal name used as subject	ˆa - entry element ˆb - part of the name
607	R	Geographical name as used as subject	
610	R	Uncontrolled subject term	
675	R	UDC	ˆa - number ˆv - edition ˆz - language of edition
700	NR	Personal name - prim. int. responsibility	ˆa - entry element ˆb - part of name

Tag	Field type: R/NR	Field description	Subfield and description
701	R	Personal name - alternative int. responsibility	ˆa - entry element ˆb - part of name
702	R	Personal name - secondary int. responsibility	ˆa - entry element ˆb - part of name
710	NR	Corporate body name - prim. int. responsibility	ˆa - entry element ˆc- addition to name or qualifier ˆe - location of meeting ˆf - date of meeting
922	NR	Type of document	

Legend: NR=non-repeatable; R=repeatable

2 REVIEW OF PUBLISHED RESEARCH

A comprehensive review of the literature is made to identify some research on data conversion in order to compare the methods and experiences with libraries that have had the same experience.

If a library developed its own method of organising the bibliographic information, it would isolate the library. With the advent of new information technology and its application in the library and information field, a number of bibliographic databases are being created worldwide and used across the national and international boundaries through networks.

The creation of a standardised bibliographic database for sharing and exchanging of data, contained in the files of the database, through networks has become a common feature among the libraries.

Standardisation of the format and structure of databases is necessary to facilitate sharing and exchanging data in an efficient and effective way. The use of MARC standard allows libraries to better share bibliographic resources, and also to replace one system with another with assurance that the data will still be compatible (Das, 2004).

The record structure of UNIMARC is defined in the ISO 2709 standard. The internal structure of CDS/ISIS is also based on this standard, so the conversion between the internal format and the exchange format can be achieved efficiently. Hopkinson (1997), describing the WINISIS programme, explains all the features of CDS/ISIS format. He also emphasises that this format is easily convertible to other formats because it is based on the UNIMARC format.

One method of generating a data map is to create a spread sheet-style table that compares the fields of one format with the fields of another format. According to Phillips (1999), this method effectively maps the data of both formats. The transfer of data from fields of the old system to the correct fields in a new system requires a data map (Phillips, 1999).

The author assumes that in all conversions a major issue is data cleanliness or records quality. He discusses that most vendors prefer the database be cleaned of errors before the conversion because the new system operates properly only with quality data. Also, the librarians are more familiar with the

old system, which helps them find errors, than with the new one. However, the old system has limited reporting and data visualisation capabilities. Phillips (1999) concludes that only a well-managed data conversion project can meet the library requirements.

For the preparation of mapping tables and its specification, the *MARC Standards Office Library of Congress* has published the document *"UNIMARC to MARC 21 Conversion Specifications Version 3.0 (August 2001)*. This document provides specifications for the conversion of bibliographic records in UNIMARC to MARC 21. A mapping table contains two main columns, one for UNIMARC and its fields and subfields and one for MARC 21. The last column is intended for the conversion notes. The decision to *Keep, Do not convert, Change,* or other decisions, and instructions for the conversion of bibliographic records from one format to another are specified in detail.

Institutions have developed different bibliographic standards for creating and developing machine-readable library catalogue records. These variations in standards and the transfer of data from one system to another has become a significant problem. Chandrakar (2002) describes how to convert bibliographic catalogue records from CCF to MARC 21 for institutions in India.

After describing the two formats to be converted, the author presents three basic components of records in order for the exchange programs to succeed: *1. Physical Structure. 2. Content Designators. 3. Content of the Records.* After analysing these components, the author presents the table of mapping formats and the problems during mapping.

Once the problems are recognised due to the mapping table, they solve the problems with a "definition" presented in the mapping table. This "definition" is the implementation note on how to convert the data automatically (Chandrakar, 2002).

Khurshid and Kadry (2006) demonstrate the experiences of data migration of the KFUPM library in Saudi Arabia. They describe a detailed analysis of DOBIS/LIBIS data structure, data extraction, including data mining, and data conversion of bibliographic, non-bibliographic and bilingual data carried out to provide maximum information to libraries about the complexities of source data and how to overcome them.

The method by the authors is the mapping data of both formats in the mapping table. The major data problems came from the source (DOBIS/LIBIS database) due to the incompatibility of DMARC with MARC 21. Their results show a successful conversion rate of 98.2 per cent, even if manual work still needs to be done to improve the quality of the database.

De Keyser (2000) lists some difficulties met during the conversion process. Most of the conversion problems are related to the repeatable fields of WINISIS. Sometimes, repeatability is not compatible with other formats. The content of one source field must be divided into subfields, e.g. title and parallel title. The text strings must be replaced with codes.

When a library decides to convert its database to another library system, studied steps should follow. These steps should conclude with a successful conversion. The step-by-step approach for the migration of the existing library data to the new software are described from Matoria and Upadhyay (2005).

These steps begin with the study of the existing system (ALTTC library in Uttar Pradesh in India), the export of the existing data into the output file, the mapping data with mapping table by presenting the fields and subfields of the formats. These steps also include the modification and improvement of data through the conversion process, and the testing of the data after the conversion.

In the end of their research, they come to the conclusion that during the conversion process some lessons and experiences are learnt. During the process, the data are refined and cleaned. They also believe that an automatic conversion is much cheaper than re-keying all the data for the new software (Matoria & Upadhyay, 2005).

Walls (2011) demonstrates the circumstances, methods and outcomes of the New York University Health Sciences Libraries' (NYUHSL) migration from their previous Integrated Library System (ILS), Innovative Interfaces, Inc.'s Millennium, to the open-source ILS Koha. This experience presents the way of the conversion implementation. Before exporting data and putting it into the new system, most errors in bibliographic records are eliminated manually by two librarians. The reason is that it is easier for these librarians, even though more time-consuming, to correct the data in the old system (Millenium) than in the new one (Koha). Another reason is the lack of librarians trained for mapping the two formats and writing the conversions definitions. Then the mapping table could be easier.

The conversion of the National Library of Bulgaria database into COMARC format is another relevant experience. Karačodžukova and Tušek (2009) report their experience on the conversion of databases into the COMARC formats. The database of National Library of Bulgaria was in the CDS/ISIS format.

The authors describe the chain of the conversion procedures. The process begins with exporting data from CDS/ISIS to COMARC ongoing preparing the mapping table and the conversion algorithm to analyse the formats. In the end, they present all problems that occur during the conversion

process, such as differences in the definition of fields repeatability, differences in the presentation of intellectual responsibility, non-standard data content, correction in the text data and the lack of the cataloguing rules application.

Kelly and Nelms (2003) demonstrate the steps of migrating data from an old system to a new one, and mention several main issues: 1) the use of old and new systems for data comparisons, 2) determination of differences between the two systems, 3) dealing with issues of data changed during the migration process, and 4) the accuracy of cut-off parameters that determine which data should be included in the migration.

WINISIS and COMARC are formats which are based on the UNIMARC format. This facilitates the conversion and enables higher compatibility of the two formats; the presence of UNIMARC makes the conversion efficient. In the conversion of a database into other library systems, format standardisation and database structure are key components (Das, 2004).

When following the steps, mentioned by Matoria and Upadhyay (2005), the WINISIS data is exported to the test COBISS/Cataloguing software. Then the mapping table with fields and subfields of the formats is prepared. Once the records are analysed, data is modified, improved through the conversion process, and tested after the conversion. Karačodžukova and Tušek (2009) report a similar experience. The authors describe the conversion as the chain of procedures, from exporting data and preparing the mapping table to analysing the records.

A data map is required to transfer data from the old system into a new one (Phillips, 1999). The mapping table is used to compare the fields/subfields of the WINISIS format with the fields/subfields of COMARC/B format. Phillips (1999) discusses the quality of the records when data is cleaned before the conversion. The old system, however, is limited in reporting the errors, and the whole procedure is time-consuming. Being aware that we have knowledge to build the mapping table and define the conversion algorithm, we decided to check and clean the data after the conversion, which saved a lot of time. On the other hand, Walls (2011) presents their experience with data conversion. Despite the fact that correcting data in the old system was time-consuming, they eliminated manually most of the errors in the bibliographic records as they were not staffed by skilled librarians to map the two formats and write the conversions rules.

To be able to prepare the specifications of the mapping table, we consulted the MARC Standards Office Library of Congress webpage. Following the recommendations, we built the mapping table for both formats, COMARC/B and WINISIS.

Chandrakar (2002) describes the basic components (physical structure and the content of the records) for analysing the records of both formats during the conversion process. On the basis of this, we analysed the structure of WINISIS and COMARC/B. First, we established the differences and

similarities between the two formats, and then we analysed the content of the records. It was De Keyser (2000) who mentioned the difficulties related to the repeatability of fields, and in some cases, the (field/subfield) repeatability in WINISIS was not compatible with that in COMARC/B. Such an analysis is important for defining the conversion algorithm to convert correct data.

Another issue in the conversion process is the quality of data of the converted records. After the conversion, software controls were used to check the validity of the records via the automatic validation within the COBISS3/Cataloguing module (COBISS3/Cataloguing User Manual, p. 197). It is possible to check if the record contains errors which can be detected automatically. Through the automatic validation of mandatory fields and MARC coding, records that need reviewing can be marked, and reports generated to help identify errors (Hudson 1984). The need to automate quality control includes minimising the time spent on the database maintenance, making error detection more consistent and accurate, and saving money (Chapman & Massey, 2002).

Philips (1999) shows that records may contain errors or other anomalies even as they reside in the old system. During the conversion process, the harmonisation of the metadata of the two formats is essential; the harmonisation being the process that enables consistency across metadata standards (St. Pierre, 1999). Some features of common metadata properties include the unique identifier (e.g. field name), whether or not metadata element is mandatory or optional, etc.

Poll and Boekhorst (1996) define quality as fitness for purpose, where the purpose of a library is reader satisfaction. As to the quality in cataloguing, they define it as satisfaction of reader's needs for a catalogue that is comprehensive, current and usable, but not necessary detailed or even perfectly accurate.

The two most commonly used measures of cataloguing quality are: the "level", i.e. the amount of data contained in records, or claimed to be contained, and the "error rate", i.e. the number of errors, or types of error, found per record. The error rates can be determined against particular standards, for example the data that should be present in a "full level" record for a particular resource where the full level is defined ultimately by the cataloguing agency (Chapman & Massey, 2002).

3 CONVERSION METHOD

3.1 DATA CONVERSION PROCESS

For the purpose of this study, the COBISS2/Cataloguing software test environment is installed. The ASAL catalogue records are exported from WINISIS to COBISS2/Cataloguing, keeping the original format.

The next step is the preparation of the mapping table where the WINISIS format and COMARC/B format are presented. After fields and subfields of the formats were examined, we establish the differences between both formats. For the purpose of this study, we export the local database records to the COBISS3/Cataloguing software test environment. In doing this, statistics on the fields and subfields of the records are generated. These statistics are used in data analysis.

Then the analysis of the actual records in WINISIS and the ideal record created according to the standards (COMARC/B Cataloguing manual, ISBD standard for monographs and AACR2 rules) is made. At this stage, based on the sample of records, we determine how many fields/subfields will be improved during the conversion process. The way how to determine which sample of records to choose, is described in Section 3.2.3.

The fourth and the last step of the process is the measurement of the converted records quality. The records are validated automatically by using the software control from COBISS3/Cataloguing software module. The numbers and the types of errors detected by the program are established, and the results commented. The comparative analysis of the converted records and those of other conventional systems (COBISS.Net, WorldCat or Library of Congress) is also used. We choose a set of test records in order to determine the quality of the converted records.

3.2 MAPPING TABLE

Data mapping allows systems to effectively convert data from one metadata standard to another. A mapping table shows equivalent elements (or "fields") in more than one database schema. It maps the elements in one format to the equivalent elements in another format.

The methods with the mapping table provide data that help us obtain better understanding of both formats. The strength of this method is the ability to focus on problems (with fields, subfields, repeatability, indicators, etc.) we face during the conversion process.

The mapping table provides a powerful analytic solution that explains complex data in a simple and easy-to-understand manner.

By means of the mapping method, the records are analysed and tables prepared for mapping the local WINISIS format into the COMARC/B format. This mapping shows how many fields and subfields are converted. Based on this feedback, we changed (improved) the data through algorithm. A detailed analysis of all exported data and algorithms for the conversion of fields and subfields is carried out (*Appendix A)*. (UNIMARC to MARC 21 Conversion Specifications, 2001).

We encounter some difficulties with this method since, through the conversion algorithm used in the mapping table, we are not able to correct and eliminate automatically all the errors presented in the data at the required level. Moreover, to use this method, good knowledge of bibliographic formats (MARC, COMARC), the ISBD (M) standard and Anglo-American Cataloguing Rules (AACR2) is required. The same applies to the training of the librarians who will be included in the conversion of bibliographic records and have to be trained on how to prepare data mapping, how to interpret data from the statistical tool, how to explain the conception of the conversion algorithm, literacy, etc.

The mapping table we prepare consists of two main columns. One column contains the fields, subfields, code and data meaning of COMARC format. In the other, the WINISIS format is presented with the same elements. Yet another column is used to show the status of the fields. In this way, we compare both formats to see if the fields either comply with each other or differ from one another. In order to formulate the right definition, we use the ISBD (M) standard and COMARC/B User Manual (*Table 3.1)*.

Table 3.1 *The structure of the mapping table*

COMARC/B				WINISIS FORMAT/B				Status	NOTES
Field	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		

For each field, subfields are defined if the data status is *keep, added, changed* or *not converted* (table legend). If there are fields/subfields converted directly in the process of conversion, we call them with status *Keep (K)*. The content of the data presented in these fields/subfields in the WINISIS format is the same after the conversion in the COMARC/B format.

Example:

WINISIS format   *610a Historia*

COMARC format   *610a Historia*

If the field or subfield is presented in the COMARC format, but it is not presented in the WINISIS format, we consider it having the status *Add (A)*. After the conversion, we add this field or subfields to the ASAL database.

Example:

In WINISIS format the records has no subfield 001a (Record status).

After the conversion into the COMARC format is added by default the code n (new record)

If a field or subfield suffers modification during the conversion, it will be considered having the *Change (C)* status. An example of such status is the field or subfield segmentation based on semantic rules and separator signs ("=", ")", ":" etc.).

Example:

WINISIS format   **200a** *Historia e Shqiperise=The Albanian History*

COMARC format   **200a** *Historia e Shqiperise* **d***The Albanian history*

If for some reasons certain fields or subfields are not converted into COMARC/B format, they have the status *Not Converted (NC)*.

Since fields have subfields and these subfields are of different number and different status within the field, we analyse the subfield status because one subfield has only one status. It is very unlikely that a field would be presented with more than one status.

For the purpose of analysing data of the two formats, the problems are classified according to the status of subfields converted to the COMARC/B format. On the basis of the results, it is determined what percentage of subfields is converted directly *Keep (K)*, *Not converted (NC)*, *Add (A)* and *Changed (C)*.

The subfield analysis allows us to determine the differences between the two standards because each criterion brings a new innovation. Furthermore, the study of the subfields shows specific and sporadic problems and the way to solve them during the conversion process.

3.2.1 Auxiliary tables

For the purpose of the conversion, for adding or changing data into fields and subfields of our database, we prepared, along with the main table *(Appendix A),* also *auxiliary tables* to function of the conversion algorithm. When, depending on a specific problem, the conversion algorithm does not suffice for an automatic conversion, these tables can help. We prepared the following auxiliary tables:

- *Country of the publication code (Appendix C).* In the local database, the country publication (102a) subfield is missing. According to ISBD and the COMARC rules, the content of this field is related to the publication place (210a) subfield in WINISIS. In this case, we mapped the data from the subfield 210a to the subfield 102a through this table.

  Example:

  The data presented in subfield 210a is ***Tirane***. The adequate data mapped from the subfield 210a to the subfield 102a should be ***alb***.

- *Language code (Appendix D).* This table contains the list of data in the form of a text in WINISIS, but should be converted to a code in COMARC/B (according the COMARC rules).

  Example:

  Language of the item presented in WINISIS   ***shqipe***

  In COMARC/B it should be converted into the code   ***alb***

- *Fields not converted (Appendix E).* Fields and subfields that are not converted into COMARC/B are presented in a separate table (to make the conversion easier) and not in the mapping table.

- *Definition for NSB/NSE (Appendix F).* When a title begins with a definite or indefinite article (such as e.g. A, An, The, Das, Der, Des, Die, Il, De, La, Le, Les, Un), we put *NSB* (Non-sorting begin) "≠" at the beginning of the title and *NSE* (Non-sorting end) "≠" just before the second word in the title. In such a case, the article has no effect on sorting and searching by title.

Example:

200 0_ aThe reality effect in the writing of history

200 0_ a≠The ≠reality effect in the writing of history

3.2.2 Statistical tools

When the WINISIS data are exported into COBISS, the statistics are generated automatically in COBISS. These statistics show the number of records for each field/subfield before and after conversion *(see Appendix B),* the frequency of fields/ subfields, and whether fields/subfields are repeatable or not. They also show the maximum and the minimum number of appearing subfields for each field, as well as an example with ID number for each field/subfield *(Table 3.2).*

Table 3.2 *Statistics database of the fields/subfields before conversion*

Tag	S	Freq	MinR	MaxR	MinT	MaxT	Uniq	IdR
105		34	1	1			34	1459
	a	34	1	1	1	1	34	1459

Table 3.3 *Statistics database of the fields/subfields after conversion*

Tag	S	Freq	MinR	MaxR	MinT	MaxT	Uniq	IdR
105		2362	1	1			2362	1536001520
	a	1504	1	1	1	1	1504	1536001520
	e	1276	1	1	1	1	1276	1536002288

Legend:
Tag   - Field code (three-digit code)
S   - Subfield code (one-digit code)
Freq   - Occurrences of fields/subfields in the processed records
MinR   - The minimum number of occurrences of fields/subfields in a single record
MaxR   - The maximum number of occurrences of fields/subfields in a single record
MinT   - The minimum number of occurrences of subfields within a field in the record
MaxT   - The maximum number of occurrences of subfields within a field in the record
Uniq   - The number of records in which a field/subfield occurs
IdR   - The record in which a field/subfield is repeated several times

This table shows that field (Tag) *105 – Textual material, monographic* has one subfield 105a (S), which occurs 34 times in the processed records (Freq). The subfield is repeated maximally 1 time in one record (MaxT shows repeatability). The unique number of records in which subfields occur is 34 (out of total 6,838). The ID number (1459) is presented in the last column of the table as an example that the field contains this subfield.

After the conversion, another table of statistics is generated (*Table 3.3*) where we can check the number of fields/subfields and see whether data is lost or added. This table shows a higher number (1504) of subfields 105a (Illustration code). It also shows that, as the result of the conversion process, data is added in this subfield. After the conversion, a new subfield 105e (Index indicator) is added with 1276 records. From the table, we can establish that the repetition of subfields is the same before and after the conversion.

Even with the manuals available for the WINISIS and COMARC formats, the COBISS statistics serves as a valuable verification tool and as performance criteria before and after the conversion. In iteration processes, it helps us check and compare the fields/subfields converted to the new format, and also greatly enhances our efficiency in reading numbers and analysing data. We can check and compare the number of fields/subfields before and after the conversion. The information on the repeatability of fields/subfields is also provided through this statistics, which allows the conversion algorithm for changing and improving data to be defined.

3.2.3   Comparative analysis of actual records with the ideal record

To determine to what degree bibliographic records would be improved, we compared the ideal record with actual records in the test environment. The ideal record is a record that is in accordance with the requirements of the ISBD/M standard, the COMARC/B format and the second level of bibliographic description according to the AACR2 (Rule 1.0D2) cataloguing rules. We expect to create full level bibliographic records that meet these requirements. Examples of full level records that we used were taken from the COMARC/B User Manual (COBISS3/Cataloguing User Manual, 2009, p. 197).

After the analysis and the comparison of the records, we were able to decide what conversion rule to write in the mapping table. We used 60 records (1%) to perform the comparative analysis since it was established that adding more records would not show other types of errors. We made a decision on the number of records by using an incremental method, as follows:

1. We started the analysis with 10 records.

2. For each field, we analysed all the records in order to find the similarities and differences between the two standards; hence providing new conversion rules.

3. If the new conversion rules or corrections were provided, we incremented the number of records by 10, repeated step 2, or stopped.

3.2.4   Conversion algorithm

For each field and subfield of both formats, the mapping table also contains the conversion rules column. In the conversion process, this column is used for all algorithms. An algorithm is a short and strict definition to tell the programmer what to do with each field and subfield in the conversion process (Thomale, 2010). The conversion algorithm is used when something that is wrong is repeated. Minor, rare or isolated mistakes are better manually.

Example:

[Add code **n** as default to all records]

or

[Parse the whole data according to the existing punctuation. Delete blanks from the beginning and from the end of string]

3.3   QUALITY MEASUREMENT OF THE CONVERTED BIBLIOGRAPHIC RECORDS

3.3.1 Automatic verification of the record quality

To measure the quality of the converted records, software controls are used to check the validity of the records via the automatic validation within the COBISS3/Cataloguing module (COBISS3/Cataloguing User Manual, p. 197). Through the validation process, the number and the type of the errors can be established and then the results analysed.

A sample of 60 (1%) converted records was selected from the total number of records (6,838) in the database. We consider that 60 records, if chosen wisely, are representative enough to raise any issue that may occur. The set of these records, chosen for testing, was different from the set chosen for providing the conversion algorithm (Paragraph *3.2.3*). Using the same sample of records for the two tasks would not make sense because all the issues found in the development sample were resolved in the conversion algorithm; here we wanted to test if there were any other issues left. The testing sample consisted of monograph records. Stratified sampling was used in order to check the record quality for the specific elements of the bibliographic description. The following criteria were used for the selection of the records:

- records with only one author (A1)

- records with two authors (A2)

- records with three authors (A3)

- records described by the title heading

The criteria are the elements of the bibliographic description which represent major access points to meet the user needs. We used them to generate various situations in cataloguing in order to capture all error types. In doing this, the quality of data would be reflected in error gravity and distribution.

For this purpose, a table (*Table 3.4)* was prepared with errors for each criterion. It contained the criteria used for sampling and the number of errors for each criterion.

To determine the nature of errors occurred in the sampling, we have prepared another table that showed what type of error occurred and in which subfield (*Table 3.5*).

Table 3.4 *Measuring data quality*

No. of	Records described	Records	Total	Total

errors	by authors			described by title	Records	Errors
	A1	A2	A3			
5						
4						
3						
2						
1						
No errors						

Table 3.5 *The type of errors*

The affected field	Error type	Occurrences

3.3.2   Method for the comparison between the converted records and records from the LC and WorldCat databases

30 converted bibliographic records were compared with the same records from other databases (Library of Congress and WorldCat).The records used for evaluation were chosen randomly. We wanted to determine how many records had more/equal/less number of fields if compared to the records from other databases. For this purpose, *Table 3.6* was prepared where for each database (ASAL, LC and WorldCat) the number of records for each field according to the relevant block was presented. The *Table 3.7* shows the meaning of fields presented in these databases.

Table 3.6 *No. of fields presented in the ASAL, LC, WorldCat*

Blocks	0XX	1XX	2XX	3XX	5XX	6XX	7XX

Fields	001	010	100	101	102	105	200	205	210	215	225	300	320	500	513	600	606	607	610	675	700	701	702	712
ASAL																								
LC																								
Worldcat																								

Table 3.7 *Fields existing in the ASAL, LC, and WorldCat records*

Fields	Meaning of fields
001	Record label
010	International standard book number (ISBN)
100	General processing data
101	Language of the item
102	Country of publication or production
105	Textual material monographic
200	Title and statement of responsibility
205	Edition statement
210	Publication, distribution etc.
215	Physical description
225	Series
300	General note
320	Internal bibliographies/Indexes note
500	Uniform title
513	Added title page
600	Personal name used as subject
606	Topical name used as subject
607	Geographical name used as subject

610	Uncontrolled subject terms
675	Universal decimal classification (UDC)
700	Personal name - Primary intellectual responsibility
701	Personal name - Alternative intellectual responsibility
702	Personal name - Secondary intellectual responsibility
712	Corporate body name- Secondary intellectual responsibility

4 CONVERSION RESULTS

4.1   CONVERSION TYPES

Considering that one converted subfield can only have one status (*keep*, *not converted*, *added* or *changed*), we presented an overview of the number of the subfields for each status. Also presented were the number and the percentage (relative to the number of subfields in COBISS standard) of subfields that have one of these statuses *(Table 4.1)*. This data showed specific and sporadic problems, and how much work was done during the conversion process. The analysis of each status in the conversion was made and issues discussed, as shown below:

Table 4.1 *Subfields status during the conversion in percentage*

The status of the subfields	*No. of the subfields*	*%*
Subfields converted directly (K)	*28*	*41*
Subfields not converted (NC)	*23*	*33*
Changed subfields (C)	*17*	*26*
Added subfields (A)	*23*	*33*

Note. Added subfields (33%) are new subfields created as the result of the conversion. 33% is relative to the total number of subfields in the WINISIS format.

4.1.1   WINISIS subfields directly converted into COMARC/B format

From the mapping table, we established which subfields were converted directly (without any changes) to COMARC/B *(Appendix A)*. In the table, these subfields had the status *Keep (K)* under the Status column. According to this status (K), we determined that only 28 subfields from the total of 68 subfields (in WINISIS) were converted directly to COMARC/B *(see Table 4.1)*, which means that 41% of the WINISIS subfields were converted directly to COMARC/B.

The first reason why these subfields were converted directly to COMARC/B was the compatibility of the subfields with the WINISIS format. The second reason was the compliance of these subfields' data with the ISBDs bibliographic description and the AACR2 cataloguing Rules. In some cases, even if we found errors made by librarians, it was not possible to correct them through the conversion process (e.g. subfield 225a). We, therefore, had to convert them directly. After the conversion, they will be corrected manually.

The percentage of subfields converted directly into COMARC/B is low if the fact that both formats are based on UNIMARC/B is taken into consideration. On one hand, this is due to the fact that there is a limited number of fields and subfields in WINISIS; on the other, the other reasons are lack of cataloguing rules and errors in cataloguing monographs. All these problems are explained in the sections below.

As regards the terminology in terms of the meaning and the content of fields and subfields, we were referred to the COMARC/B format. The subfields converted directly to the COMARC/B format are presented below:

Subfield 101f – Language of title page (***R***) is part of field 101 – Language of the Item (NR) and contains the language of title page when it differs from the language or languages of the text. This subfield is repeatable for each language of the title page.

Subfield 106a – Physical medium designator (NR) is part of field 106 – Textual material physical attributes (NR) and contains coded data relating to the form of textual material. In WINISIS, only subfield *106a – regular print* exists.

From field 200 – Title and Statement of Responsibility (NR), only two subfields were converted directly from the local database to COMARC/B:

Subfield 200h – Number of part (***R***) is the section for items which are identified by common name and subsection name. All records (1187) in WINISIS that contain subfield 200h are converted to COMARC/B *(Appendix B)*.

Subfield 200i –Name of part (R) is the section for items which are identified by common name and subsection name. In WINISIS, 550 records with subfield 200i exist, and the same number remains the same after converting data to COMARC/B *(Appendix B)*.

Subfield 205a – Edition statement (NR) from field 205 – Edition Statement (NR) corresponds to the ISBD Edition Area. The subfield 205a/NR contains a word, phrase or group of characters in a formal statement, identifying an item as a member of an edition. There are 396 records that contain *subfield 205a/NR*; after the conversion, the number of the records remains the same in COMARC/B in the respective *subfield 205a/NR (Appendix B).*

From field 215 – Physical description (NR), all subfields were converted to COMARC/B. This field contains information on the physical characteristics of the item and corresponds to the ISBD Physical Description Area.

Subfield 215a – Specific material designation (NR) contains the number of pages of the monographs. The numbers of records that contain subfield 215a/NR in WINISIS corresponded with the number of records in subfield 215a/NR in COMARC/B.

Subfield 215c – Other physical details (NR) contains data about details of the illustrative matter. From the statistics before and after conversion, the same number of records was noticed.

Subfield 215d – Dimensions (NR). The linear measurement of an item and/or dimension relevant to the use of the item. Before conversion the number of records that contained subfield 215d was 5501. And the number of subfields remained the same also after the conversion (5501).

Subfield 215e – Accompanying material (NR). It contains a brief description of any material accompanying the item being described. The table of statistics showed 25 records containing subfield 215e before and after the conversion.

Other subfields added from the local database to COMARC/B were subfields from field 225 – Series (/R). This field contains the title of the series along with any other title information and statements of responsibility relating to the title including any of the preceding repeated in other languages in the form and sequence in which they appear on the item being catalogued. It corresponds to the ISBD Series Area.

Subfield 225a – Series title (NR), *subfield 225v – Volume of designation (R)* and *subfield 225x – ISSN of series (R).* There are many errors in these subfields that had been made by librarians. These errors are very complex and it was impossible to correct them automatically. This is why we converted this data without changing it and kept the original data from the local format.

Example:

225 a Systeme International d'information pour les sciences et la

technologie agricoles;ISSN 10-10-3325

In this example, the ISSN number is recorded in subfield 225a instead of subfield 225x/R.

Subfield 330a – Text of note (NR). A note containing summary, abstract or short contents of the item presented in the language coded in subfield *330z – Language*. The number of records remained the same before and after the conversion.

Subfield 600a – Entry element (R) and *subfield 600b – Part of name (R)* are converted directly to COMARC/B. These subfields are part of the field 600 – Personal name used as subject (R). This repeatable field contains the name of a person who is one of the subjects of the bibliographic item. We converted successfully all 200 records containing subfield 600a and 156 records containing subfield 600b.

Subfield 607a – Entry element (R) is part of the field 607 – Geographical name used as subject (R). This repeatable field contains a geographical name used as a subject heading of the bibliographic item. This field contains only one subfield *600a – Entry element*. On the basis of the statistics, it was established that the number of the records (696) remained the same before and after the conversion *(Appendix B)*.

Subfield 610a – Subject term (R) is converted directly to other format, and contains uncontrolled subject. This repeatable field contains subject terms that are not derived from controlled subject heading lists and that are not covered by fields 600–609 concerning their contents and use. Even if during cataloguing process some mistakes were noticed in this subfield, they could not be corrected automatically. We converted these data as presented in the original database.

  Example: in WINISIS   610aLiterature aAlbania

        instead of   610aAlbanian literature

Subfield 675a – Universal decimal classification (NR). This subfield contains a class number applied to the item according to the Universal Decimal Classification scheme, with an indication of the edition being used. On the basis of the statistics, we provided the same number of records (6821) that contained subfield 675a before and after the conversion. The result was satisfactory with 100% converted data.

From field 700 – Personal name – Primary intellectual responsibility (NR) we converted the following subfields directly to COMARC/B:

Subfield 700a – Entry element (NR). The portion of the name used as the entry element in the heading that enables both search and sort within the catalogue. It is mostly the surname of the author. If the first part of the name contains two or more elements, they are all entered in subfield a. If field 700 is entered a subfield a may also be entered. On the basis of the statistics, the result was the same before and after the conversion.

Subfield 700b – Part of name (not entry element) (NR). The remainder of the name (forename or names), used when the entry element is a surname of family name.

We converted these subfields directly even if they contained mistakes. These mistakes were made by librarians who used the initials of the author's name: e.g. Kadare, I. instead of Kadare, Ismail. In some cases, the author's name was not written in authorised form: e.g. Cvajg, Stefan instead of Zweig, Stefan.

Subfield 701a – Entry element (NR) and *subfield 701b – Part of name (NR)* are part of field 701 – Personal name – Alternative intellectual responsibility (R). This field contains the name of a person considered to have alternative intellectual responsibility for a work described in the bibliographic record. The form of the name appearing as the uniform heading is always entered in field 701. We converted all the records with subfield 701a (628 records) and subfield 701b (620 records).

Field 702 – Personal name – Secondary intellectual responsibility (R). This field contains the name of a person consider to have secondary intellectual responsibility for a work described in the bibliographic record (e.g. illustrator, translator, editor, mentor, photographer, etc.). The form of the name appearing as the uniform heading is always entered in field 702. Directly converted were *subfield 702a – Entry element (NR)* and *subfield 702b – Part of name (NR)*. The number of records with subfield 702a (2934) and subfield 702b (2926) was same before and after the conversion.

Field 710 – Corporate body name – Primary intellectual responsibility (NR). This field contains the name of the corporate body considered to have primary intellectual responsibility for a work described in the bibliographic record. The form of the name appearing as the uniform heading is always entered in field 710. Directly converted were *subfield 710a – Entry element (NR)*, *subfield 710e – Location meeting (R)* and *subfield 710f – Date of meeting (NR)*. In the original database, we had only 5 records with subfield 705a, and the result was same after the conversion, which means that no data was lost during the conversion process.

4.1.2   WINISIS subfields not converted into COMARC/B format

The percentage of subfields not converted to COMARC/B is presented in *Table 4.1*. There are altogether 23 subfields (out of total 68 subfields) from the WINISIS database that were not converted to COMARC/B. The analysis shows that the percentage (33%) is high if compared to the total number of subfields presented in WINISIS. This happened due to irrelevant data presented in these subfields. Another reason was the fact that some subfields contained data related to holdings data (copy number, number of place and location), which should be converted to the COMARC/H format.

For subfields with status *not converted (NC)*, we used an auxiliary table *(Appendix E)*. According to the reason why these subfields were not converted from WINISIS to COMARC/B, the subfields were classified into three groups:

1. *Subfields that exist only in some records with unknown and irrelevant data*

2. *Subfields that do not exist in COMARC/B and will be converted to COMARC/H*

3. *The content of these subfields was already added by default to other subfields in all records.*

1. *Subfields that exist only in some records with unknown and irrelevant data*

These subfields contain unknown and irrelevant data that concerns the aim of these subfields. The data was not converted to COMARC/B.

Not converted subfields are as listed below:

Table 4.2 *Subfields that exist only in some records with unknown and irrelevant data*

No.	Field	Subfield	Meaning of fields and subfields
1	021/R	a, b	Legal deposit number
2	022/R	b	Government publication number
3	100/NR	a	General processing data
4	102/NR	a	Country of publication or production
5	410/NR	1	Series
6	421/NR	a	Supplement
7	422/R	a	Parent of supplement
8	510/R	a	Parallel title proper
9	512/R	a	Parallel title
10	601/R	a, e, x, 2	Corporate body name used as subject

11	675/R	v, z	*Universal decimal classification (UDC)*
12	712/NR	a, b, 4	*Corporate body name – secondary intellectual responsibility*

Some of this irrelevant data exist due to errors made by cataloguers *(see example below, subfield 4101/NR)*. This data is not related to the content of the item. While the title gives us data about the tradition culture in *Europe in the modern time*, subfield 4101/NR shows data about *Balance of payments statistics yearbook*. This means that this data is no longer important. We decided not to convert the subfield because we could clean the database during the conversion process.

Example: ID=1463

```
010  a99927-615-0-4  d700 L.
100  a19890601d1989    km y0engy0103    ba
101  amul - Multiple languages  ffre - French
102  aDE -
105  aa   z   101yy -
106  ar - regular print
200  aKultura popullore në Evropë në fillimet e kohës së re
210  aPa vd.  cCEU;Arbri  d1996
215  a381 f.  cme il.  d21 cm.
301  a39 B 959
302  aBASH
410  1200 (TI=Balance of Payments Statistics Yearbook. AU=Carson, Carol
     S.)  aIFLA publications
607  aEvropa
610  aKultura popullore - Evropa
675  a39(4)  vBASH  zalb
700  aBurke  bPeter
702  aQesteri  bSerzh  4Përktheu nga anglishtja -
712  aIFLA Section of Art Libraries  4400 - funder/sponsor
801  aPT  bBN
922  am
```

2. *Subfields that do not exist in COMARC/B and will be converted to COMARC/H*

These three subfields (copy number, identification number and location) contain data related to the COBISS3/Holdings module *(COMARC/H format) (Table 4.3)*. In WININIS, this type of subfields is in the same software module, whereas in COBISS, they belong to the *Holdings* module.

Table 4.3 *Subfields that do not exist in COMARC/B and will be converted to COMARC/H*

No.	Field	Subfield	Meaning of fields and subfields
1	006/R	a	Copy number
2	301/R	a	Note pertaining to identification number
3	302/R	a	Location

The example below shows the WINISIS data which should be converted to the COMARC/H format.

Example: ID=1466

```
1466. ID=1466 LN=0000001466 N   V1  23.04.2012 EKMB::LIBRARY

006  a50297
006  a50298
010  a99927-45-36-3  d500 L.
100  a19890801d1983    k  f0engy0103    ba
101  aeng - English
102  aGB -
105  ay   z   101yy -
106  ar - regular print
200  aNjë miqësi evropiane
210  aT.  cOnufri  d2001.
215  a206 f.  c21 cm.
301  a821.18 K 635
302  aBASH
600  aFaik Konica
601  aInternational Cataloguing-in-Publication Meeting,  eOttawa,  f198
     x[Proceedings]  2SIPOR
606  aCataloguing in publication  y[Meetings]  2SIPOR
607  aShqipëria
610  aLetërsia shqipe-shkrime periodike
More...
```

3. *The content of these subfields was already added by default to other subfields in all records*

This group consists of *subfields the content of which was added by default to other subfields* in all records. This is shown in the table below.

Table 4.4 *Subfields will be added by default to other subfields in all records*

No.	Field	Subfield	Meaning of fields and subfields
1	801/R	a	Originating source
2	922/R	a	Type of material

In field ***801/R***, fixed data ($aPT $bBN) is presented, and is not relevant to the aim of the subfield. Data of the field is automatically added to subfield 000. Field ***922/R*** exists only in 50 records and contains code for bibliographic level *m – monographs*. By default, the code is added to subfield 001c of all records *(see Example)*.

Example: ID=1458

```
1458. ID=1458 LN=0000001458 N   V1  23.04.2012 EKMB::LIBRARY

006  a50283
006  a50284
100  a19860401d1986    k  y0engy0103    ba
101  apor - Portuguese  ceng - English
102  aPT -
200  aUdhëtari dhe hija e tij
210  aT.  cBargjini  d2001
215  a282 f.  d20 cm.
301  a159.9 N 649
302  aBASH
304  aTransl. of: A dictionary of the European Communities
610  aUdhëtari
610  aLetërsia gjermane
610  aMaksima
610  aPsikologji
675  a159.9.019.3
675  a821.112.2-84
700  aNietzsche  bFriedrich
702  aHatia  bTaulant  4Përktheu nga origjinali -
More...
```

4.1.3   Subfields added during the conversion process

When studying the mapping table, we gave all subfields the status *Add (A) (Appendix A)*. To add new subfields, we used the conversion algorithm. Proportionally to the total number of the WINISIS subfields (68), we added 23 subfields to the COMARC/B format. Thus after the conversion, 33% of the WINISIS database contains newly added subfields *(Table 4.1)*.

We had to add some subfields which are obligatory according to the COMARC/B standard. Additionally, some subfields were added to complete and improve the database with useful and new data. The conversion process allowed us to add semi-automatically new subfields by means of the conversion algorithm that defined added data by default, using either auxiliary tables (prepared before conversion) or important data from another fields presented in the WINISIS database (*see Chapter 3*).

With these methods we analysed all subfields added to COMARC/B. All the problems and difficulties we faced in the conversion concerning *Added* fields are presented in this thesis.

To be able to efficiently convert the database, we had to study both formats. We also had to know which fields and subfields are obligatory and which should have a bibliographic description according to the required standards. Also important is to know which fields and subfields help library users in searching for information, and which fields and subfields are useful for cataloguers.

In the light of all mentioned above and according to the COMARC/B standard for monographs, ISBD and the AACR2 Cataloguing Rules, we improved the database, as following:

Field 001 – Record label (NR) is part of *0XX – Identification Block* (COMARC/B User Manual). This block contains data that identify the record or the item recorded in it. This field contains a general data required for cataloguing records and it is mandatory. It does not exist in WINISIS and was, therefore, added to all records of the WINISIS database. Although this field is not visible in OPAC to users, it is very useful for cataloguers and completes the database with more information. The following subfields are mandatory in COMARC/B format *(Table. 4.5)*.

Table 4.5 *Mapping table for field Record label (001/NR)*

COMARC/B				Status	NOTES
Field	Subfield	Code	Meaning of subfield		
001/NR			**RECORD LABEL**	A	
	a/NR		Record status		
		n		A	Add code **n** as default to all records.
	b/NR		Type of record		
		a		A	Add code **a** as default to all records.
	c/NR		Bibliographic level		
		m		A	Add code **m** as default to all records.
	d/NR		Hierarchical level code		
		0		A	Add code **0** as default to all records.
	7/NR		Script of cataloguing		
		ba		A	Add code **ba** as default to all records.

These subfields were added by using the algorithm: "*Add code… as default to all records*".

In *subfield 001a – Record status (NR)* one of the character codes indicates a record cataloguing status. It was added by default to all records the code *n (new record)*.

Subfield 001b – Type of record (NR) has the code *a (textual material, printed)* as default to all records because the material in WINISIS is textual material, printed.

In *subfield 001c – Bibliographic level (NR)* code *m (monograph*) was added by default. This means that a bibliographic entry is complete in one part, or intended to be completed within a finite number of parts. All records in WINISIS are monograph records, which is why the code *m* was added.

In *subfield 001d – Hierarchical level code (NR)* code *0 (no hierarchical relationship)* was added by default.

In *subfield 0017– Script of cataloguing (NR)* code *ba (Latin)* was added by default because the language of cataloguing is Albanian and Albanian is the Latin script.

Example of the field 001/R after the conversion

001 an - regjistrim i ri **b**a - material tekst, i shtypur **c**m - monografi **d**0 - nuk ka lidhje hierarkike e000041 **7**ba - latin

Field 100 – General Processing Data (NR) *from Block 1XX (Coded Information).* This field does not exist in WINISIS. In COMARC, it contains coded data applicable to records. After analysing the data, we decided to add the following subfields *(Appendix A)*:

> *b – Type of publication date*
> ***c – Publication date 1***
> *d – Publication date 2*
> *e – Target audience*
> *g – Modified record code*
> ***h – Language of cataloguing***
> *i – Transliteration code*
> ***l – Script of title proper***

The subfields in bold are mandatory in COMARC/B. Through the conversion algorithm method, the codes and data for each of the above subfields were defined.

Subfield 100b – Type of publication date (NR). The code indicates the type of date entered in subfield *100c – Publication date 1* and subfield *d – Publication date 2*. This subfield is related with data in subfield *210d – Year of publication*. Depending on the data in subfield 210d, the code for *subfield 100b* should be *d – monograph complete when issued, or issue within one calendar year* or *g – monographs, whose publication continues for more than one year.*

From the total number of records (6838) after the conversion, there were only 6673 records where subfield 100b was present *(Appendix B)*. The subfield was not added to all records because subfield 210d is not presented in all records in WINISIS.

Subfield 100d – Publication date 1 (NR) and *subfield 100c – publication date 2 (NR)* are related directly to data from *subfield 210d. Subfield 100c* is mandatory in COMARC/B when cataloguing monographs. According to the conversion algorithm, these subfields were filled with data taken from *subfield 210d*.

Example:

100 **bd** - monograph complete when issued, or issued within one calendar year **c2001** em - adult, general **g**0 - unmodified record **h**alb - Albanian **i**y - no transliteration scheme used **l**ba - Latin

210 **a**Tiranë **c**Dritëro **d2001**

Subfield 100e – Target audience (NR) is defined, in general, by default with code

m – adult because ASAL is a special library in function of a special target group of users.

Table 4.6 *Mapping table for field General processing data (100/NR)*

COMARC/B				WINISIS FORMAT/B			Status	NOTES
Field	Subfield	Code	Meaning of subfield	Field	Subfield	Meaning of subfield		
100/NR			**GENERAL PROCESSING DATA**	210/R		**PUBLICATION**		Convert only data from the first field 210 Data from field 210 is used for conversion in fields: 100, 102, 210. After the conversion into field 100, keep field 210 for following conversions.
	b/NR		Type of publication date		**d/R**	Date of publication	A	• Convert data only from the first subfield 200d.
		g **d**						• Do not convert data that starts with letters
	c/NR **d/NR**		Publication date 1 Publication date 2				A	• Delete blanks from the beginning and from the end of string.
								• If subfield 210d contains 4 numbers at the beginning of the string, following by "-" (without spaces before and after "-") and then again 4 numbers, add code **g** in subfield 100b.
								• Convert first 4 numbers into subfield 100c and 4 numbers after "-" into subfield 100d.
								• Otherwise add code **d** in subfield b and convert only first 4 numbers from subfield 210d into subfield 100c.
	e/NR		Target audience code				A	

COMARC/B				WINISIS FORMAT/B			Status	NOTES
Field	Subfield	Code	Meaning of subfield	Field	Subfield	Meaning of subfield		
		m						Add code **m** as default

Subfield 100g – Modified record code (NR). This code indicates whether the character set available within the system is sufficient to transcribe the data as found on the item. For this purpose, the auxiliary table of the Item's language was prepared *(Appendix D)*.

The table contains the *Language of the item* column. Depending on the language, the adequate column for *script of the item*, *modification code* and *transliteration code* is prepared. This table helped us define data in subfields 100g, 100h, 100i, 100l.

To define the modification code, subfield *101a – Language of the Item* was used. Depending on the language of the Item, the code was defined.

> *Example:*
>
> **100 b**d - monograph complete when issued, or issued within one calendar year **c**2001**e**m - adult, general **g**0 - unmodified record **halb Albanian i**y - no transliteration scheme used **l**ba - latin
>
> **101**0 **a**alb - **Albanian**

Subfield 100h – Language of cataloguing (NR). This subfield is mandatory and is defined by default with code *alb – Albanian Language* because Albanian is the language of cataloguing in ASLA.

Subfield 100i – Transliteration code (NR). This code indicates the type of transliteration used. In relation to the language of the item, adequate data was added to subfield 100i. The codes used were *y – no transliteration scheme used* when the script of the item is Latin, or *b – other transliterations* when the script of the item is Cyrillic.

Subfield 100l – Script of title proper (NR). This code indicates the script of the title proper. The data is used to define a display script, which is why the code entry is mandatory. This code is added from subfield *101a – Language of the Item*. According to the language of the item, the code should be *ba – Latin script* or *ca – Cyrillic (Appendix D)*.

From field 102 – Country of publication or production (NR), we added **subfield 102a – Country of publication**. This field contains codes for one or more countries of publication or production of the item. Field 102a contains a three-character code representing the country in which the item was published. The codes are to be entered according to ISO 3166.

To complete subfield 102a, data from subfield *210a – Place of publication* is used. A table is prepared where each place of publication is defined by the code of country *(Appendix C)*.

According to the statistics, only 6034 records from all 6838 records of database (82.2%) had the code of the country before and after the conversion *(Appendix B)*. In the remaining records, subfield 102a was empty and remained empty after the conversion. The reason is that not all the records have data in subfield 210a.

In field 105 – Textual Material, Monographic (NR) subfields were added, as follows. This field contains coded data relating to monographic textual material. Through the algorithm *"if 215c exits, add code **a** in the **subfield 105a"**,* the code *a –illustrations* was added to subfield *105a*. Subfield *215c – other physical details* contains data about the illustrations. After the conversion, the code *a* was added to 1503 records (22%) *(Appendix B)*.

Also **subfield 105e – index indicator** was added. Data presented in subfield *300a –Text* note or *215d – Dimensions* were used according to the algorithm. The code *1 –index present* was used in this subfield for all records containing the word *index* in subfield *300a* or *215d*. 1276 converted records will have this code (18.6%) *(Appendix B)*.

*Example***:** Field 105/R added after the conversion

105 aa - illustrations e1 - index present

From field 200 – Title and Statement of Responsibility (NR) the subfields were added as follows:

This field corresponds to the ISBD Title and Statement of Responsibility Area. This is an important area in the bibliographic description because in this area data is written as it is found in the title page. Through the conversion process, missing data were added according to ISBD (M) standard.

Subfield 200f – First statement of responsibility (R) and **subfield 200g – Subsequent statement of responsibility (R)** do not exist in WINISIS. The data about Statement of Responsibility is presented in fields 700, 701, 702, and 710. With the use of the algorithm method, this data was added as following:

Example:

700 _1 a Antoine b Serge

701 11 a Vilmorin bJean-Baptiste de

701 11 a Yana bAndré

200 0_ f Serge Antoine, Jean-Baptiste de Vilmorin, André Yana

Field 330 – Summary or Abstract (R) contains a summary, abstract or a short description of the item contents. In the field 330/R, we added ***subfield 330z – Language***. This field is a three-character code of the language of a summary or abstract. Here, we added by default the *code alb – Albanian*

Field 610 – Uncontrolled Subject Terms (R) is a repeatable field that contains subject terms that are not derived from controlled subject heading lists and that are not covered by fields 600-609 concerning their contents and use. ***Subfield 610z – Language (NR)*** with code *alb* (Albanian) was added by default to all records that contained the subfield subject term (610a).

Example:

6100 a Polikondensantë të sintetizuar **zalb** - Albanian

6100 a Prodhime sintetike **zalb** – Albanian

Field 675 – Universal decimal classification (R) contains a class number applied to the item according to the Universal Decimal Classification scheme, with an indication of the edition being used. ***Subfield 675c – UDC access (NR)*** contains the UDC number appearing as a uniform class number within the COBISS system. It is used for either COBIB shared database or local databases retrieval. This subfield is mandatory. We added the temporary *code "fik"* by default to all records. In the second stage, once the conversion is concluded, we will prepare a controlled list with UDC scheme and we will manually add the UDC code.

Field 700 – Personal Name – Primary Intellectual Responsibility (NR). This field contains the name of the person considered to have primary intellectual responsibility for a work described in bibliographic record. The form of the name appearing as the uniform heading is always entered in field 700. ***Subfield 7004 – Relator code (R)***. The code is entered to designate the relationship between the person named in the field and the bibliographic item to which the record refers. The code *070 – Author* was added by default.

Field 701 – Personal name – alternative intellectual responsibility (R) contains the name of a person considered to have alternative intellectual responsibility for a work described in the bibliographic record. The form of the name appearing as the uniform heading is always entered in field 701. In field 701 **subfield 4 – Relator code** may be entered strictly. The code *070 – Author* was added by default.

4.1.4   Subfields changed during the conversion process

To determine what percentage of subfields we changed during the conversion process, we analysed the mapping table *(Appendix A)*. From the table, we took into account all subfields with the status *Change (C)*. Out of the total number of subfields existing in WINISIS (68 subfields), changes were made to 17 subfields (26%). The percentage and the graphic for this proportion are presented in *Table 4.1*.

Also analysed was the record quality. For this purpose, the method of comparing the ideal record with actual records in the test environment was used. The ideal record was created in accordance with the requirements of the ISBD standard, the COMARC format and the second level of bibliographic description according to the AACR2 Cataloguing Rules.

With the use of the conversion algorithm, bibliographic description of the records was improved automatically. Not all changes done in this conversion improved the database in the required level. Below are the results analysed and interpreted in detail:

Field 010 – International Standard Book Number (NR). ISBN is a unique numeric commercial book identifier. The field corresponds to the ISBD Standard Number (or Alternative) and Terms of Availability Area. In the world of book publishing, the ISBN is one of the most important numbers will be encounter. ISBN number is a very important search element in OPAC. Users can easily find a book in OPAC if they know its ISBN number. There is a compliance of both formats as regards field 010 (in all subfields a, b, d). The ISBN number should be recorded correctly according to the rules of the ISBN agency.

In **subfield 010a – ISBN (NR)**, the number should be used with hyphens. There are some errors presented in subfield 010a:

- There was a punctuation mark ″.″, but no hyphens between the ISBN numbers.

- Some records contained the acronym ISBN. According to the rules, the letters ISBN, which are usually printed along with an ISBN on the item, should not be entered in subfield 010a. In the original database, this mistake was corrected manually because this was a complex problem that could not be corrected automatically through the conversion algorithm. Also, this mistake was found in 10 records.

- The qualification data was entered in subfield 010a instead of subfield 010b. Through the conversion algorithm, data in subfield 010a was corrected automatically.

Example: ID 23

Before conversion:

010 a2-02-000668-5(Tome 8)

After conversion:

010 a2-02-000668-5   bTome 8

Table 4.7 *Changed data in field 010*

010/R	--			INTERNATIONAL STANDARD BOOK NUMBER	010/R			INTERNATIONAL STANDARD BOOK NUMBER				Once converted data from field 010 is not used any more.
	a/NR b/NR			ISBN number		a/R		ISBN number			C	Parse the subfield according to the brake "(".
993/R	--											
		a/R									C	• Parse the whole data according to the existing punctuation. Delete blanks from the beginning and from the end of string. • Convert the first part of the parsed string into subfield 010a, but only up to the first letter that is different from letter "x" or "X". • Convert the part of the parsed data that is after "(" into subfield 010b. Do not convert punctuation "(" and/or ")". • If the subfield do not contain "(", convert whole string into 010a.
010/R		b/NR		Qualification		b/NR		Qualification			C	If there is any data in subfield add space and then convert data from subfield b. Do not convert "(" and/or ")"

Subfield 010b – Qualification (NR). An identification of the scope of the ISBN in subfield a (if present): usually the name of a publisher, an indication of the binding of the item, or an indication of the relationship of an ISBN to a set or to a particular volume. Within the COBISS software, qualification is entered without brackets because they are displayed automatically. Through the conversion algorithm, the brackets were deleted in subfield 010b.

Example:

Before conversion:

010 a5-03-001166-8 **b**(russk)

After conversion:

010 a5-03-001166-8 **b**russk

There are mistakes made due to different structure of the WINISIS program, and mistakes made by librarians who do not apply the relevant standards during cataloguing.

Despite the use of the conversion algorithm, errors still remain in the subfield 010a. Some of the errors are corrected by hand in the original database of the WINISIS program. Others will be corrected after the conversion.

Subfield 101a – Language of text (R). This subfield contains the code of the language text of the item. Unlike the COMARC/B format, in the WINISIS format subfield 101a contains the language of the text in the word form. Also, in WINISIS subfield 101a is not repeatable; repeatable is field 101a (*see example*).

For this purpose, we prepared an auxiliary table *(Appendix D)* that for each data (the language text) in WINISIS contained the adequate data (the code of the language text) in COMARC/B (Table). Through the conversion algorithm, we changed data in subfield 101a/R (*see Appendix D*). After the conversion, 2963 records (44%) were provided with subfield 101a/R *(see Appendix B)*.

Example:

Before conversion:

101 **a**fre - French

101 **a**spa - Spanish

101 **a**por - Portuguese

After conversion:

101 afre-French **a**spa-Spanish **a**por-Portuguese

Table 4.8 *Language code*

WINISIS Language	COMARC/B Code
eng	eng
frëge	fre
frënge	fre
frëngjisht	fre
ger	ger
gjermane	ger
gjuha	alb
greke	gre
hungareze	hun

maqedonase	*mac*
maqedone	*mac*
polake	*pol*
polonisht	*pol*
portugeze	*por*
rumune	*rum*
ruse	*rus*

Subfield 101c – Language of intermediate text (R) contains the language of any intermediate translation when the item is not translated from original. In WINISIS, subfield 101c contained by error in some records the same data as in subfield 101a. Using the conversion algorithm, we deleted data as shows the example below:

Example:

Before conversion:

101 ashqipe - **c**shqipe

After conversion:

101 ashqipe

Field 200 – Title and statement of responsibility (NR). This field contains the title along with any other title information and statements of responsibility relating to the title including any of the preceding repeated in other languages (parallel titles, parallel statements of responsibility, etc.). It corresponds to the ISBD Title and Statement of Responsibility Area.

When a title begins with definite or indefinite article (or some other word having no filing value), the article followed by a space is given between two signs NSB/NSE "≠". The article has no filing value in sorting and searching *(see Appendix F)*. According to the COMARC/B format standard, this definition was added in field 200/NR.

Example:

200 0_ **a**The reality effect in the writing of history

200   **a**≠The ≠reality effect in the writing of history

Subfield 200a – Title proper (R). The chief title of the item includes alternative title but excludes other title information (e.g. subtitles) and parallel titles. A subfield is mandatory for every record. The WINISIS format does not have subfield 200d/R. In such a case when cataloguing monographs, the parallel title is entered in field 200a, using the punctuation "=" between *(see example)*.

Example:

200a Medical dictionary=Dictionaire médicale=Medizinisches

Wörterbuch=Dizionario medico

With the conversion algorithm, we corrected subfield 200a as follows:

2000 a Medical dictionary **d**Dictionaire médicale **d**Medizinisches Wörterbuch **d**Dizionario medico

As a result, 220 converted records (31%) contain subfields 200d *(see Appendix B)*.

Subfield 200e – Other title information (R). Subtitles and other title information that appear subordinate to the title proper in subfield a, c or d. Repeatable for each segment of other title information and for parallels other title information.

This subfield exists in WINISIS, but it happens that sometimes in some records two or three subtitles are entered by error in the same subfield 200e, separated with colon ":".

Example:

Before conversion:

200^e Zhvillimi ekonomik dhe impakti ambiental: Seminar italo-shqiptar Tiranë, 17-18 tetor 2000 : në bashkëpunim me Univ."La Sapienza" të Romës

After conversion:
2000 a Administrimi i gjeoresurseve **e**zhvillimi ekonomik dhe impakti ambiental **e**seminar italo-shqiptar Tiranë,17-18 tetor 2000 **e**në bashkëpunim me Univ."La Sapienza" të Romës

Out of 3358 records that contained subfield 200e, we created in each of 3408 records a repeatable subfield 200e/R, beginning with small letters according to ISBD standards.

Field 210 – Publication, Distribution, etc. (NR). This field contains information on the publication, distribution and manufacture of the item including associated dates. It corresponds to the ISBD Publication, Distribution Area.

According to the ISBD standard and COMARC standard, field 210/NR is not repeatable. In WINISIS, however, this field is repeatable. With the use of the conversion algorithm, we could not resolve the problem due to the complexity of other problems that followed. So, we decided to use field 993

(field for locally defined content) that contains data which cannot be entered into other specific field or subfield within COMARC/B. It can be saved only locally, and cannot be found in COBIB.

In records with repeatable fields 210, we converted only the first field 210, whereas other repeatable fields 210 were converted to field 993. In the second stage of the conversion, we can correct these records manually once the conversion process is concluded.

Example:

Before conversion:

210^a Zagreb**^c**Illirycum**^d**1991

210^a Zagreb**^c**Rilindja**^d**1991

After conversion:

210 aZagreb **c**Illirycum **d**1991

993 **a**Zagreb ; Rilindja ; 1991

Subfield 210a – Place of publication, distribution, etc. (NR). The town or other locality where the item is published or distributed entered in the form it is quoted on the item. Repeatable for each place named.

In some records there is more than one place of publication separated with semicolon ";" punctuation. With the use of the conversion algorithm, we converted the second place of publication in another repeated subfield 210a/R.

Example:

Before conversion:

210^a Montreal; Paris**^c**Décarie; Masson**^d**1991

After conversion:

210 a Montreal **a**Paris **c**Décarie **c**Masson **d**1991

In a number of records, in subfield 210a, as the place of publication "T." is entered instead of "Tiranë". According to the rules of the bibliographic description, this is not correct. With the use of the conversion algorithm, this was corrected automatically.

Example:

Before conversion:

210^aT.^cDritëro^d1997

After conversion:

210 a Tiranë **c**Dritëro **d**1997

Subfield 210c – Name of publisher, distributor, etc. (R). The name of the publisher or distributor which may be in a short form so long as it is readily recognisable (*see the example above*).

Subfield 300a – Text of note (R). The general note entered in any form. In this subfield, data related to Bibliography notes or Index note is entered. According to the standards, this specific data should be entered into **subfield *320a – Internal Bibliographies/Indexes/Summaries Note (R)***. This subfield contains a note indicating that the item contains a bibliography, index or summary. With the conversion algorithm we changed the subfield as shown below:

Example:

Before the conversion:

300 a Bibliografi f.274-288

After the conversion:

320 a Bibliografi f.274-288

The example below shows that the correction of the data was partial. With the conversion algorithm we could convert this data from subfield 300a/R to subfield 320a; however, we could not convert the data *"Indeks"* to another subfield 320a/R. This kind of problems will be solved manually after the conversion.

Example:

Before conversion:

300 a Bibliogr. në tekst. Indeks.

After conversion:

320 a Bibliogr. në tekst. Indeks.

The fields listed below were converted automatically into the field 300a/R. After checking the data in these fields, we noticed that they contain general data about notes and not data required for these fields.

Subfield 304a – Note Pertaining to Title and Statement of Responsibility (R). This subfield contains a note relating to the title or statement of responsibility as entered in field 200.

Subfield 305a – Note Pertaining to Edition and Bibliographic History (R). This subfield contains a note relating to the edition of the item or to its bibliographic history.

Subfield 312a – Note Pertaining to Intellectual Responsibility (R). This subfield does not exist in COMARC/B. With the conversion algorithm we converted it to field 300/R.

Subfield 327a – Contents Note (R). This subfield contains a note describing the contents of the item.

Field 702 – Personal Name – Secondary Intellectual Responsibility (R) contains the name of a person considered to have secondary intellectual responsibility for a work described in the bibliographic record (e.g. illustrator, translator, editor, mentor, photographer, etc.). The form of the name appearing as the uniform heading is always entered into field 702. The Authority control is implemented in this field. Here, the data should be very strictly.

Also **subfield 7024 – Relator code** must be entered strictly. In WINISIS, subfield 7024/R contains data expressed with words. In COMARC/B, there is a list with codes that are related to subfield 702a/R. To correct this subfield according to the standard, we prepared a list with all data existing in WINISIS and the adequate code in COMARC/B.

Example:

```
bashkepun*    070
bashkëpun*    070
perkth*       730
përkth*       730
```

pergat*	220
përgat*	220
shqiper*	730
shqipër*	730
edit*	340
ured*	340
redak*	340
red.	340
ed.	340
traduz*	730
transla*	730

Before *conversion:*

702 **a**Altman **b**Franz-Lothar **4**red.

After *conversion:*

702 1 **a**Altman **b**Franz-Lothar **4**340-redaktor

4.2 QUALITY MEASUREMENT OF THE CONVERTED BIBLIOGRAPHIC RECORDS

4.2.1 Automatic verification of record quality

The sample of 60 converted records was analysed based on automatic validation available within the COBISS3/Cataloguing software. We classified records by the criterion of "one author" (A1), "two authors" (A2), "three authors" (A3) and by title. In the table, the number of errors for each criterion is presented *(Table 4.9)*.

Table 4.9 *Measurement of data quality*

No. of errors	Records described by authors			Records described by title	Total Records	Errors
	A1	A2	A3			
5	1	2	-	4	7	7x5=35
4	-	1	1	3	5	5x4=20
3	3	2	4	2	11	11x3=33
2	7	6	7	5	25	25x2=50
1	-	1	1	1	3	3x1=3
No errors	4	3	2	-	9	Total errors=141
Total errors	28	33	31	49		

Table 4.10 *Type of errors*

The affected field/subfield	Error type	Occurrences
010a	Wrong ISBN format number	5

100l	Missing script of text	54
100c	Publication date 1 is empty	1
101a	Language of text is empty	46
102a	Missing place of publication	4
200 (ind)	Indicator 1 required	13
210c	Missing name of publisher	8
210d	Missing Date of publication	3
210a	Missing Place of publication	1
70x4	Missing the kind of Relator code	6

According to the statistics, the highest number of errors is in subdivision "Records described by title" (49 errors). As shown in Table 4.9, we see that the majority of errors suffer from *the indicator* missing in field 200 (Title). Namely, in COMARC/B, the indicator 1 should be in field 200 if the record is described by title (which means the title is important as it functions as the heading of the record). In WINISIS, records do not have indicators.

According to Table 4.10, the highest number of errors occurs in *subfield 100l – Script of text* and in subfield *101a – Language of text* (70%). With the use of the validation programme in COBISS3/Cataloguing software, we can establish that in the records these subfields are missing. In COMARC/B they are obligatory, in WINISIS they do not exist. During the conversion process, we were able to add these subfields only to 3119 records (out of 6838 records) because of missing data in some records of the original database. As a result, these errors are expected and are inherent. However, these can also be seen as the limitations which must be overcome either algorithmically, by entering the default value, or manually. If these errors were excluded, the number of errors would decrease significantly.

The subdivision with least errors is A1. This group of records does not have any special type of errors as it contains inherent errors.

Another error that occurred in all subdivisions is the missing *Relator code* in field 70X *Personal Name – Intellectual Responsibility*. During the conversion process, we added by default to field *700 Personal Name – Primary Intellectual Responsibility* and to field *701 Personal Name – Alternative Intellectual Responsibility* the author's code "070" in all records; in field *702 Personal Name – Second Intellectual Responsibility* of the original database, data is entered in subfield 7024 only in 2338 records. In the rest of the records, there are no data to define the role of the second author in the record.

The format of an ISBN is another type of error. It occurred in 8% of the records. The automatic conversion detected errors in the field format, e.g. the wrong format of ISBN numbers. This error is the result of the incorrect format entered into the WINISIS records during the cataloguing process.

Errors in subfields *210a – Place of publication*, *210c – Name of publisher*, and *210d – Date of publication* are due to missing data in these subfields in the WINISIS database. As a result, data is missing also in subfields *100c – Publication date 1* and *102a – Country of publication* because according to the COMARC/B rules, subfield 100c is related to subfield 210d and subfield 102a to subfield 102a.

The statistics also showed that 61% of records had up to 2 errors and 39% more than 3 errors. These errors include format errors, missing fields and incorrect entries of data (the WINISIS database). The majority of errors occurred under "Records described by title" due to missing indicator data, the least under A1.

Our conclusion is that records can be improved further. One of possible solutions is to improve the algorithm and thus eliminate errors due to language (subfield 101a) and errors due to script text (subfield 1001). The sheer size of these errors is more than 70% of the errors found in the sample records. However, there are ways available to reduce these record data limitations. If they could be reduced to a minimum, this would greatly enhance the quality of records to better comply with the COMARC/B format rules.

Records with errors due to missing data/field or wrong entries can be improved also. However, wrong data entries, which are randomised errors in statistical sense, cannot be much improved. The records statistics shows that the percentage of the randomised errors is lower than that of non-randomised errors. Hence, there is considerable potential for improvement in the records data by using algorithmic techniques and supplementing the missing data according to the ISBD and COMARC B standards, and the AACR2 cataloguing Rules.

4.2.2 Comparison between the converted records and records from the LC and WorldCat databases

A random sample of 30 converted bibliographic records was compared with the same records from the Library of Congress (LC) database. We checked how many fields are in each record of the ASAL database, the LC database and the WorldCat database *(Table 4.11)*.

Table 4.11 *No. of fields presented in the ASAL, LC, WorldCat*

Blocks	0XX		1XX				2XX					3XX		5XX		6XX					7XX			
Fields	001	010	100	101	102	105	200	205	210	215	225	300	320	500	513	600	606	607	610	675	700	701	702	712
ASAL	30	30	30	30	28	12	30	2	30	30	17	5	18	-	-	3	-	13	30	30	30	4	14	-
LC	30	30	30	30	26	23	30	3	30	30	8	17	20	3	5	8	25	11	-	-	29	-	10	4
Worldcat	30	29	30	30	26	23	30	2	30	30	7	16	20	3	5	8	25	11	-	-	29	-	10	3

The table illustrates that records in all three databases have almost the same number of fields. One of the differences lies in the distribution of the number of these fields. If we analyse fields from *5XX – Related title block*, we can notice that no records of ASAL database contain fields 500 and 513 (Table 4.11). The field 500 contains the uniform title, whereas the field 513 contains data on added title page. These fields contain the particular title selected by a bibliographic agency by which a work that has appeared under varying titles is to be identified for cataloguing purposes. In the original WINISIS database, these fields do not exist.

Another difference can be found in *6XX – Subject analysis block*. Among the converted records, there are no records containing field *606 – Topical name used as subject*. This is because in the LC and WorldCat databases controlled subject terms (LCSH) are used in subject analysis; in the ASAL database uncontrolled terms in the classification of the subject are used. In COMARC, field 610 is used for uncontrolled subject terms. This is the reason why field 610 is present in all 30 of the converted records.

The statistics shows that only in 3 converted records, when compared with 8 records of LC and WorldCat), field *600 – Personal name used as subject* contains data about the name of the person who is one of the subjects of the bibliographic item. This field is essential for users. As an access point in OPAC, it enables users to search data by keyword. So ASAL should improve its records by adding to them also this piece of information.

It is evident from the statistics that field *675 – Universal Decimal Classification* is contained in all 30 records from the ASAL database, but not in the LC records; LC, namely, uses Dewey Decimal Classification (DDC).

In the ASAL records, also field *712 – Corporate body name – Secondary intellectual responsibility for the publication* exists but is not used by cataloguers. The lack of the information in this field makes a record less informative. One of the reasons why the field is missing in WINISIS records is the lack of standards in the Albanian language, which prevents cataloguers to define a corporate body.

We established that field *300 – General note* from *3XX – Notes block* cannot always be found in the ASAL records as it is not obligatory. It, however, contains additional data for all fields in bibliographic format.

Field *105 – Textual material monographic* exists in the LC records, but not in ASAL records. Through the conversion, we added this field only to some records (see Section 4.3), which means that this was done only in 50% of ASAL records.

4.3   DISCUSSION

After mapping the data of the WINISIS format and COMARC/B format, we analysed the similarities and differences between the two systems in terms of the structure of the catalogue records.

We determined that out of 68 subfields presented in the WINISIS format there were only 28 (41%) subfields converted directly into the COMARC/B format with status K (Keep). The directly converted subfields are the result of the compatibility of the fields/subfields structure of the formats studied. Considering that both formats are based on the UNIMARC format, we expected a higher number of subfields to be converted directly into the other format. The low percentage is a result of the differences in field repeatability which causes the data not to convert according to the COMARC/B format rules (see Section 4.1.3). Missing data in some fields and subfields causes the records not to meet the requirements of the ISBD/M standard and

COMARC/B standard. Non-bibliographic data presented in some subfields and the mistakes made by cataloguers during the cataloguing process are the reasons that we do not have more subfields with status K (Keep).

Another possible status of the conversion is NC (not converted). The results show that 23 subfields or 33% (out of 68 subfields), presented in WINISIS, were not converted into the COMARC/B format. This number is very high compared to the total number of subfields in the local format. This can partially be explained with the subfields containing irrelevant data for the aim of the field. Another reason is the different structure of the WINISIS format and COMARC/B format. In the WINISIS format, the fields that contain the administrative data (e.g. copy number, shelf mark and location number) are presented in the same segment of the bibliographic description, while in the COMARC/B format this data should be in another segment (COMARC/H format).

Consequently, some of the subfields were filled with the default value in all the records, which is another reason for non-converted data. Even if this percentage is high, we did not lose any important data, as the data was transferred from these subfields in the WINISIS format to other subfields in the COMARC/B format. The irrelevant data that was not converted did not have any value.

We have changed 26% of the subfields of the WINISIS format (17 subfields out of 68 subfields present in the WINISIS format). The data was changed based on the following criteria: (i) data that was put in the wrong fields was corrected, (ii) repeated fields/subfields were corrected in accordance with the COMARC/B format, (iii) textual data in the WINISIS format was converted to coded data in the COMARC/B format, (iv) some of the fields were divided into more subfields in order to fit the COMARC/B format, (v) some of the mistakes made by cataloguers who recorded the bibliographic data incorrectly were corrected. Due to the NSB/NSE characters, the access points (title, series, publication etc.) are more accessible for the library users.

Karačodžukova & Tušek, (2009) found similar problems in their conversion from WINISIS format to COMARC/B format: differences in the definition of field repeatability, nonstandard data content for registered printers and publishers in the national bibliography for which they had to establish a new field at the local level. Another problem they had to solve was the corrections in the text data, where a list of rules was prepared at the time of the conversion. The problem of inconsistent application of cataloguing rules was solved by using an algorithm.

As a result of the conversion, we increased the fields/subfields of the WINISIS database by 33% (status A, added 23 more subfields to 68, the total number of records in the WINISIS format). This is a considerable percentage. One of the main reasons is that the WINISIS format contains a limited

number of fields/subfields. Another reason is that the local librarians have used a minimum level of the bibliographic description. We should also emphasise that some of the fields/subfields in the COMARC/B format are obligatory, hence a record cannot be considered to be correct without them.

On the other hand, the conversion allows us to add some data which can be very helpful to library users. This data was missing from the WINISIS format and by use of automatic methods (algorithms) we added and improved the data from the local format.

The comparison of the actual records with the ideal record provided the correction of the database according to the ISBD/M areas, the second level of the AACR2 rules and the COMARC/B format standard. We have to consider the subfields with status C (Change) with 26% and A (Add) with 23% to determine how much the local database has improved. The results show that after the conversion the WINISIS database is improved by 59% compared to the original database before the conversion. Even if some mistakes remained after the conversion, we are satisfied with this result.

Khurshid & Kadry (2005) in their study on data conversion came to a satisfactory result of 98.2% even if some of the problems due to the incompatibility of the DMARC format with MARC 21 remained. Using the mapping method they were able to find a solution and achieve good results. Compared to them, we obtained a conversion accuracy of 100%.

The quality of the converted data is an important issue in the conversion process. Statistically, there were 141 errors in 60 records. There was a variety of errors: format errors, missing fields and incorrect data from the input records (the WINISIS database). Cataloguing errors also include information that is incorrect, uncontrolled etc. Such errors are also defined for the most part in a set of standards such as AACR2, COMARC/B format or ISBD (M) standards.

Error typologies indicate an order of importance. For example, Chapman and Massey (2002) began with "major" and "minor" errors. There are records with one error but that does not automatically mean that it is a "minor" error, because it can occur in important fields which can be an access point of the record. On the other hand, there are records with numerous errors in non-mandatory fields which make the database less complete in terms of information, but do not make it inaccessible.

For a catalogue to be an effective tool, it must have some ''qualities''. Samad (1994) mentions some of the quality characteristics of a catalogue record, such as: error free, especially in their access points, adherence to nationally accepted cataloguing rules and standards (AACR, 2nd ed., LC Subject Heading, or any standard classification scheme).

Manaf and Rahman (2005) found an average of almost 2.5 errors per record, but the majorities were deviations from AACR2 or LC rules. From our statistics, we found an average of 1.28 errors per record, mostly due to AACR2 rules and incorrect input.

Database quality consists of the aspects of accuracy, completeness, consistency and currency according to the theoretical discussion of Fox, Levitin and Redman (1994). From the comparison of our converted records with the records from the LC and WorldCat databases, we can say from our results that almost 90% of the converted records are complete with data in their fields, even if these records contains some errors as mentioned above.

The conversion process has its own cost, such as hours of studies, programming time, debugging time, etc. But if we compare the manual conversion with the process of re-cataloguing the records in the new format, we can say that the automatic conversion is much cheaper in terms of time and money. Matoria and Upadhyay (2004) in their study about data migration to another system came to the conclusion that they converted all 13,000 records successfully by using mapping methods and saved a substantial amount of money compared to the cost of re-keying all the data in the new software.

Finally, based on this study we arrive at the conclusion that the conversion was efficient. We achieved a successful conversion rate of 100%. The conversion was performed without losing any data. The applied methods helped us get to know the structure of the formats, to analyse data and to find solutions for improving and correcting data. However, we have to say that not all errors were corrected through conversion methods. We realise that some of the errors were not corrected automatically. After the automatic conversion process, the data will be corrected manually.

The records are converted to a better format (COMARC format) compared to the WINISIS format. The other libraries in Albania using the WINISIS program now have a model for converting their data into the COMARC format.

5 CONCLUSION

Libraries and information workers live in a world where the development of information, along with global competition is changing the shape and space of information services. If they do not support these changes they will stay behind, so they need help and support when developing their profession and their library services.

This concern for the development of information resources and construction of databases to suit the needs of their communities was also present in various libraries in Albania who strived to develop information resources and construct databases to fit the needs of their communities. Lack of financial resources and management policies dictated them to find information systems provided free of charge by various international organisations. This was the case in the system integration of CDS/ISIS DOS first, followed by WINISIS from UNESCO.

To reach the broadest community of users, information must be made available in accordance with a number of related metadata standards. The goal of the creation of a standardised bibliographic database for sharing and exchanging data has become a common feature among libraries. In this context, the library of the Academy of Sciences of Albania decided to update their bibliographic records by converting the data from the WINISIS format to the COMARC/B format of the COBISS software. The conversion provides online access for users allowing the library to come out of isolation. Being part of the shared cataloguing system prevents the duplication of work and allows the library to share bibliographic resources more efficiently. The shared catalogue allows the library to acquire cataloguing data that is predictable and reliable.

This thesis provides a model for converting bibliographic data from a local library format (the WINISIS format) to the COMARC/B format. To reach this goal, our study was based on some research questions. The first question was: *How compatible are the two formats?* To verify the compatibility we determined which fields and subfields of the WINISIS/B format correspond to the COMARC/B format.

We found that only 41% of the WINISIS subfields are converted directly and 33% of the subfields are not converted into the COMARC/B format. This means that even though both formats are based on the UNIMARC format, they are only partially compatible.

In the second step, we prepared a mapping table in which the WINISIS format and COMARC/B format are presented. The methods within the mapping table provided data that helped us obtain a better understanding of both formats. In the mapping table we defined the subfields without any change with the status *K (Keep)*. The subfields that were not converted into the COMARC/B format were given the status *NC (Not converted)*. The subfields added during the conversion process have the status *A (Add)*. When we made changes in the subfields, we allocated the status *C (Change)*. The statistical tools provided by the conversion programme were used in data analysis. The statistics generated by the programme show the numbers of the records for each field/subfield before and after conversion.

The other question that guided our study was: *How much will the bibliographic records improve?* To answer this question, we compared the actual records with the ideal record defined according to international standards. The ideal record meets the ISBD/M standard, the second level of the bibliographic description of the AACR2, 1.0D2 Rules, and the COMARC/B format. In the next step, we analysed the actual records in the WINISIS format by comparing them with the ideal record created according to COMARC/B Cataloguing standard, ISBD/M standard for monographs and AACR2 Cataloguing Rules. The comparative analysis was performed on a sample of 60 records (out of 6838 records in total). The number of records was chosen by using an incremental method. Findings showed that 26% of subfields were changed and 33 % were added during the conversion process. Through the comparison of the records from the local database with the ideal record we found that our records were missing some fields and subfields, they contained errors due to the different structure of the formats studied and also contained errors created by librarians in the cataloguing process.

The other research question was: *Of what quality will the converted bibliographic records be?* The converted records are verified automatically through the COBISS3/Cataloguing software. We also compared the converted bibliographic data with the records of well-established databases (WorldCat, Library of Congress). The last research question is: *How efficient is the conversion?* Our study comes to a conclusion about the efficiency of the conversion.

In the last step, we measured the quality of the converted records with the abovementioned methods. A sample of 60 records of the converted records was selected from the population. Stratified sampling was used in order to verify the quality of the records for specified elements of the bibliographic description (records with one, two, three authors and records with title heading).

We analysed the conversion results by focusing on the status of the subfields. From the mapping table we determined 28 (41%) subfields with status K (Keep). These subfields were converted directly to COMARC/B format because of the compatibility of the formats. These formats are based on the UNIMARC format, which allow the exchange of data.

The subfields with status NC (Not converted), not converted into the COMARC/B format, represent 33% (23 subfields) of the WINISIS database. This is due to the irrelevant data presented inside these subfields. The other reason is that some of the subfields contain holdings data (copy number, number of place and location) which is not presented in the COMARC/B format but in the COMARC/H format instead.

There were 23 subfields (out of 68 subfields) with status A (Add), which represents 33% of the WINISIS database. We have added new subfields because some of the subfields are obligatory in the COMARC/B standard. The subfields are added also to complete and improve the database with useful and new data.

The results show that 26% (17 records out of 68) of the WINISIS database have status C (Change). Through the conversion algorithm we have automatically improved the bibliographic description of the records. Not all the changes made in this conversion have improved the database to the required level.

The conversion of the bibliographic records from the WINISIS format to COMARC/B format was a fruitful experience with many challenges and difficulties.

In the conversion process, we analysed the features of the WINISIS format/B and COMARC/B format. Through mapping methods, we compared the two different systems and defined the problems that arose during the conversion. The study proved that the conversion is applicable and can be made efficient due to the automated process. The conversion can also be applied to other standards if the format's particularities are taken into account.

The conversion process had to deal with a major issue, namely data cleanliness or quality. The problems are the following:

- cataloguers recorded the bibliographic data incorrectly,

- missing data in some fields and subfields,

- non-bibliographic data found in fields and subfields,

- repeated fields and subfields,

- fields with irrelevant data to the aim of said field,

- inconsistent formats and special alphanumeric format,

- data in the wrong fields,

- commas and other punctuation marks.

The majority of the data problems were caused by incorrect cataloguing and human errors made by cataloguers. A lot of cleaning up work is still to be done to improve the quality of the database. However, the conversion was done without losing any data.

This study proved that it is a compulsory practice to strictly follow international standards in creating bibliographic descriptions. It is best if library software is based on the MARC format, which allows the exchange of data.

Some of the key areas for improvement of records identified by the analysis of data include the following:

- The characteristics of both formats used: both formats are based on UNIMARC format and apply the ISBD and AACR2 rules for bibliographic description.

- Some fields and subfields are incompatible and the number of fields and subfields is not the same for both formats.

Analysing the conversion methods, we arrived at these conclusions:

- The structure of the WINISIS format and the COMARC/B format are compatible for most of the fields and subfields.

- The records are improved during the conversion results. The records converted into the COMARC/B format contain more metadata than the records in the local format. The access points were expanded and cleaned. The quality of the local records was improved as the results of the conversion process.

- The records are fully converted automatically. This conversion is much cheaper and less time consuming than re-cataloguing data for the second time, making the conversion an efficient method.

The conversion process is not an easy task. The structure of the data is different among systems. From a collection of 6838 bibliographic records (for monographs) we had no loss of data during the conversion. 100% of the records in the WINISIS format were converted successfully into the COMARC/B format. We are satisfied with the conversion.

For future work, we will further improve (manually) the actual records with the ideal ones according to international standards. We have determined how convertible the WINISIS format/B is into COMARC format/B (in percentage, presented with graphics and charts). We determined the advantages and disadvantages of this conversion method; we improved the records and brought them to the required level.

The problems described above and their solutions can help other libraries convert their cataloguing records from one system to another.

After completing the conversion, the records' functionality will be checked, using the COBISS3/Cataloguing, COBISS3/Holdings and COBISS/OPAC software. The assessment of their functionality will consist in the possibility of their use in the library operations system, in the public catalogue and in the bibliographies management system.

6 REFERENCES

CDS/ISIS Reference Manual (2004). Prepared by UNESCO

http://www.unesco.org/isis/files/winisis/windows/doc/english/WINISIS15rev.pdf

Chandrakar, R. (2002). Mapping CCF to MARC21: An Experimental Approach.

Cataloging & Classification Quarterly, Vol. 33, No. 1, pp.33 -49. Retrieved from: http://www.tandfonline.com/doi/abs/10.1300/J104v33n01_05

Chandrakar, R. (2005). An approach to mapping CCF to Dublin Core. *The electronic*

library, Vol. 23, No. 5, pp. 577-599. doi: 10.1108/02640470510631290

Chapman A. & Massey O. (2002). A catalogue quality audit tool. *Library and Information*

Research News, no. 82, Spring 2002, pp. 26-37. doi: 10.1108/01435120210432282

Chapman, A. (1994). Up to standards? A study of the quality of records in a shared

cataloguing database. *Journal of librarianship and Information science*, No. 26, pp.201-210. doi: 10.1177/096100069402600404

Chudamani, K. S. & Nagarathna, H. C. (2006). Interoperability between Dublin Core, UNIMARC, MARC21, with AACR2R as the standard frameworks

for cataloguing in the digital environment. *4th Convention PLANNER-2006, Mizoram Univ., Aiyawl, 09-10 November, 2006.* Retrieved from*:*

http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/1218/185-195.pdf?sequence=1

COBISS Platform – Co-operative Online Bibliographic System and Services
http://www.cobiss.net/cobiss_platform.htm

COBISS3/Cataloguing: user manual (2012). (internal document).

COMARC/B format for bibliographic data: user manual (2008). Maribor: IZUM.

Curwen, A. G. (1997). UNIMARC and international record exchange: an overview of

recent projects and developments. *Program*, Vol. 31, No. 3, pp. 227-238.

doi: 10.1108/EUM0000000006887

Das, D. (2004). MARC 21: the standard exchange format for the 21st Century. *2nd Convention PLANNER-2004, Manipur Uni., Imphal, 4-5 November,*

2004. Retrieved from:

http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/434/04Planner_20.pdf?sequence=1

Fox, C., Levitin, A. & Redman T., "The notion of data and its quality dimensions", *Information Processing and Management*, 1994, Vol. 30, No. 1, pp. 9-

19. Retrieved from: http://www.sciencedirect.com/science/article/pii/0306457394900205

Györkös, J. & Seljak T. (2006). COBISS.NET: support to knowledge and intercultural

dialogue. *In The role of universities and the competitiveness of the Danube region* Maribor, 21 to 23 September 2006 (pp. 215-219).

Hopkinson, A. (1997): CDS/ISIS AND UNIMARC. *MARC workshop, Fourth* International *Conference Crimea 97*. Retrieved from:
http://scholar.google.com/scholar?hl=sl&q=CDS%2FISIS+AND+UNIMARC&btnG=

Ivanovic, D., Surla, D. & Konjovic Z. (2011). CERIF compatible data model based on
MARC 21 format. *The electronic library*, Vol. 29, No. 1, pp. 52-70.
doi: 10.1108/02640471111111433

Joyce, H. & Ting, Z., (1995). Migration from LS/2000 to KeyNOTIS: a medium-sized
academic library's experience. *Information Technology and Libraries.* Retrieved from:
http://connection.ebscohost.com/c/articles/9703143421/migration-from-ls-2000-keynotis-medium-sized-academic-librarys-experience

Karačodžukova, N. D. & Tušek, S. (2009). Konverzija katalogov Nacionalne knjižnice
Bolgarije v format COMARC. *Organizacija znanja*, let. 14, zv. 4. Retrieved from: http://splet02.izum.si/cobiss-oz/news.jsp?lang=svn&init=t&cptype=utf

Kelly, C & Nelms, C. (2003). Roadmap to checking data migration. *Computers &*
Security, Vol. 22, Iss. 6, pp.506-510. Retrieved from: http://www.sciencedirect.com/science/article/pii/S0167404803006084?via=ihub

Keyser, P. De (2000). Conversions to and from CDS/ISIS: an overview of the possibilities,
problems and the programs. *Program: electronic library and information systems*, Vol. 34, Iss: 2, pp.159 – 175. Accessed from:
http://www.emeraldinsight.com/journals.htm?articleid=862389

Matoria, R. M. & Upadhyay, P. K. (2005). Migration of data from one library management
system to another: a case study in India. *Program: electronic library and information systems*. Vol. 39, Iss: 2, pp.160 – 166. Retrieved from:

http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=EJ921888&ERICExtSearch_SearchType_0=no&accno=EJ921888

Poll, R. & Boekhrost P. (1996). *Measuring quality: international guidelines for performance measurement in academic libraries*. London: K.G. Saur.

Rogina, A. (2008). The use of the UNIMARC Format in COBISS.Net. *UNIMARC USERS GROUP 2nd Meeting - 14 March 2008, Italy*. Retrieved from: http://unimarc2meeting.bnportugal.pt/

Rudic, G & Surla, D. (2009). Conversion of bibliographic records to MARC 21 format. *Electronic Library*, Vol. 27 Iss: 6, pp.950 – 967. doi: 10.1108/02640470911004057

Samad, N. A (1994). Quality cataloguing to improve user access. *Proceedings of the 1994 Joint PPM/LAS Congress: Towards Achieving High Performance Libraries, Vision for Future*

Seljak, M. & Seljak. T. (2002). The development of the COBISS system and services in Slovenia. *Program: electronic library & information systems*, Vol 36, No 2, 2002 , pp. 89-98. doi: 10.1108/00330330210429316

St. Pierre, Margaret (1999). Issues in crosswalking content metadata standards. *Information Standards Quarterly*, Vol. 11, No. 1, pp. 2-6. Retrieved from: http://www.niso.org/publications/white_papers/crosswalk/

UNIMARC to MARC 21 Conversion Specifications Version 3.0 (August 2001). Prepared for the Network Development and MARC Standards Office Library of Congress. Retrieved from: http://www.loc.gov/marc/unimarctomarc21_intro.pdf

Walls, I. (2011) "Migrating from Innovative Interfaces' Millennium to Koha: The NYU Health Sciences Libraries' experiences", *OCLC Systems & Services*, Vol. 27 Iss: 1, pp.51 – 56. doi: 10.1108/10650751111106564

Yasser, Chuttur M. (2011). An Analysis of Problems in Metadata. *Journal of Library Metadata,* 11:51–62. Retrieved from: http://www.tandfonline.com/doi/abs/10.1080/19386389.2011.570654#preview

Khurshid, Z. & Kadry H. M. (2006). Data migration from DOBIS/LIBIS to Horizon: experiences from the KFUPM Library in Saudi Arabia. *Program: electronic library and information systems*, Vol. 40, Iss: 1, pp.74 –

88.

doi: [10.1108/00330330610646825](10.1108/00330330610646825)

Zuraidah A. M. & Ramlee A. R., (2006). Examining the quality of National Library of

Malaysia (NLM) cataloguing in publication (CIP) records. *Library Review*, Vol. 55, No. 6, pp. 363-373. doi: [10.1108/00242530610674776](10.1108/00242530610674776)

APPENDIX A: Mapping WINISIS format into COMARC/B format

Data base: **ACADEMY OF SCIENCES OF ALBANIA LIBRARY (Bibliographic Description of Albanian Books 2004-2011)**

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
000/NR	--			IDENTIFICATION BLOCK						
		a/NR								
		1-3		entry mask (type)					A	001
		4-5		version of the record label (01-99)					A	01
		b/NR		entry date YYYYMMDD					A	Conversion date
		c/NR		user-creator of the record					A	BASH::OLD
		e/NR		normal record					A	'0'c
		f/NR		copy date from the shared database YYYYMMDD					A	'0'
		g/NR		10-digit local number					A	The initial number
		t/NR								Flag for C3 records

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
		u/NR		Data of conversion from other systems YYYYMMDD						Data of conversion from other systems
001/NR	--			**RECORD LABEL**					A	
		a/NR		Record status						
			n						A	Add code **n** as default to all records.
		b/NR		Type of record						
			a						A	Add code **a** as default to all records.
		c/NR		Bibliographic level						
			m						A	Add code **m** as default to all records.
		d/NR		Hierarchical level code						
			0						A	Add code **0** as default to all records.
		7/NR		Script of cataloguing						
			ba						A	Add code **ba** as default to all records.
010/R	--			**INTERNATIONAL STANDARD BOOK NUMBER**	010/R			**INTERNATIONAL STANDARD BOOK NUMBER**		Once converted data from field 010 is not used any more.
		a/NR b/NR		ISBN number		a/R		ISBN number	C	Parse the subfield according to the first brake "(.
										Parse the whole data according to the existing punctuation. Delete blanks from the beginning and from the end of string.
										Convert the first part of the parsed string into subfield 010a.
										Convert the part of the parsed data that is after "(" into subfield 010b. Do not convert punctuation "(" and/or ")".
										If the subfield do not contain "(", convert whole string into 010a.
010/R		b/NR		Qualification		b/NR		Qualification	C	If there is already some data in

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										COMARC subfield 010b, add space and then convert data from subfield b. Do not convert "(" and/or ")"
010/R		d/NR		Term of availability and/or price		d/NR		Price	C	Do not convert (" and/or ")"
100/NR	--			**GENERAL PROCESSING DATA**	210/R			**PUBLICATION**		Convert only data from the first field 210. Data from field 210 is used for conversion in fields: 100, 102, 210. After the conversion into field 100, keep field 210 for following conversions.
		b/NR		Type of publication date		d/R		Date of publication	A	• Convert data only from the first subfield 200d.
			g d							• Do not convert data that starts with letters
		c/NR d/NR		Publication date 1 Publication date 2					A	• Delete blanks from the beginning and from the end of string. • If subfield 210d contains 4 numbers at the beginning of the string, following by "-" (without spaces before and after "-") and then again 4 numbers, add code **g** in subfield 100b. • Convert first 4 numbers into subfield 100c and 4 numbers after "-" into subfield 100d. • Otherwise add code **d** in subfield b and convert only first 4 numbers from subfield 210d into subfield 100c.
		e/NR		Target audience code					A	
			m							Add code **m** as default
					101/R			**LANGUAGE OF THE ITEM**		Convert data only from first field 101.

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										Data from field 101 is used for conversion in fields: 100 and 101 After the conversion into field 100, keep field 101 for following conversions
		g/NR		Modified record code		a/R		Language of text	C	Convert data only according to the first repeated subfield 101a. Use table lang_code. If the data from subfield 101a is in the column "WINISIS Language", add adequate data from column "Modification code" into the subfield 100g. Otherwise add the code **0** as default.
		h/NR		Language of cataloguing					A	
			alb							Add code **alb** as default to all records.
		i/NR		Transliteration code		a/R		Language of text	A	Convert data only according to the first repeated subfield 101a. Use table lang_code. If the data from subfield 101a is in the column "WINISIS Language", add adequate data from column "Transliteration code" into the subfield 100i. Otherwise add the code **y** as default
		l/NR		Script of title proper		a/R		Language of text	A	Convert data only according to the first repeated subfield 101a. Use table lang_code. If the data from subfield 101a is in the column "WINISIS Language", add adequate data from column "Script" into the subfield 100l. Otherwise add the code **ba** as default.
101/NR				**LANGUAGE OF THE ITEM**	**101/R**			**LANGUAGE OF**	A	After conversion field 101 into field

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
								THE ITEM		101, do not use field 101 anymore. Data from repeated fields 101 convert into repeated subfields 101a
	0_								A	If there is only subfield 101a or If there are subfields 101a and 101c but with same data.
	1_									If there are subfields 101a and 101c with different data.
		a/R		Language of text		a/R		Language of text	C K	Use table lang_code. • If the data from subfield 101a is in the column "WINISIS Language", add adequate data from column "COMARC/B Code" into the subfield 101a. • If subfield 101a does not exist or there is not data from subfield 101a in table lange_code, do not add anything.
		c/R		Language of original work		c/NR		Language of original work	C K	Use table lang_code. • Convert only data from 101c which is different from data in 101a. • If the data from subfield 101c is in the column "WINISIS Language", add adequate data from column "COMARC/B Code" into the subfield 101c. • If subfield101c does not exist or there is not data from subfield 101c in table lange_code, do not add anything.
		f/R		Language of title page		f/NR		Language of title page	K	
102/NR	--			COUNTRY OF	210/R			PUBLICATION		After the conversion into field 102,

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
				PUBLICATION OR PRODUCTION						keep field 210 for following conversions.
		a/R		Country		a/R		Place of publication	A	Use Country_publ_code • If the data from first subfield 210a is in the column "Place of publication", add adequate data from column "Country code" into the subfield 102a • If not, do not add any default value. • If after the conversion subfields 102a contain same data, leave only one subfield.
105/NR	--			**TEXTUAL MATERIAL, MONOGRAPHIC**	**215/NR**			**PHYSICAL DESCRIPTION**	A	Data from field 215 is used for conversion in fields: 105 and 215. After the conversion into field 105, keep field 215 for following conversions.
		a/R		Ilustration codes		c/NR		Other physical details	A	If 215c exits, add code **a** in subfield 105a
			a							
					300/R			**GENERAL NOTE**		
		e/NR		Index indicator		a/NR		Text of Note	C	If the word ″inde*″ present in 300a or 215d, add code **1** in subfield105e
					215/NR			**PHYSICAL DESCRIPTION**		If not, do not add any default value.
						d/R		Dimensions	C	
			1							
106/NR	--			**TEXTUAL MATERIAL-PHYSICAL ATTRIBUTES**	**106/NR**			**TEXTUAL MATERIAL-PHYSICAL ATTRIBUTES**	K	Once converted data from field 106 is not used any more
		a/NR		Physical medium designator		a/NR		Physical medium designator	K	
200/NR				**TITLE AND STATEMENT OF RESPONSIBILITY**	**200/NR**			**TITLE AND STATEMENT OF RESPONSIBILITY**		Once converted data from field 200 is not used any more
	0_									

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
		a/R **d/R** **e/R**		Title proper		**a/R**		Title proper	C K	Parse the subfield according to the colon ":" and/or equal sign "=" punctuation (they cannot be at the beginning or at the end of the subfield)
										Parse the whole data according to the existing punctuation. Delete blank from the beginning and from the end of the string. Delete full stop (".") from the end of parsed parts.
										Convert the first part of the parsed string into 200a and other parts according to the punctuation preceding the parsed data. Convert the part(s) following an equal sign "=" into one or more subfields 200d and part(s) following a colon "_:_" into one or more subfields 200e. Do not transfer the punctuation.
										The order of the converted subfields must correspond to the order of data in the original record.
										If the subfield contains no colon and/or equal sign, convert 200a into 200a.Delete full stop (".")from the end of the string.
										For adding NSB/NSE at the beginning of subfield 200a, use the definition from the doc. below Definition for using NSB/NSE
		e/R		Other title information		**e/R**		Other title information	C	Parse the subfield according to the

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										colon "_:_" punctuation. (It cannot be at the beginning or at the end of the subfield).
										Parse the whole data according to the existing punctuation. Delete blank from the beginning and from the end of the string. Delete full stop (".") from the end of parsed parts.
										Convert each part of the parsed string into repeated subfield 200$e. Do not transfer the punctuation.
										The order of the converted subfields must correspond to the order of data in the original record.
										If the subfield contains no colon, convert 200e into 200e. Delete full stop (".") from the end of parsed parts.
										After conversion always convert first letter of the string in lowercase if it is not. For adding NSB/NSE at the beginning of subfield 200e, use the definition from the doc. below Definition for using NSB/NSE
200/NR		f/R		First statement of responsibility	700/NR 701/R 702/R			Personal name – primary intellectual responsibility Pers. name alternative int. Responsibility Pers. name-secondary int. Responsibility		Data from field 700, 701 and 702 are used for conversion in fields: 200, 700, 701, 702. After the conversion into field 200, keep field 700 701 and 702 for following conversions. If there is not fields 700 and/or 701

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										in record, convert fields 702 into subfield 200f. Ignore fields 702 with more than 2 subfields. If there <u>is</u> at least one field 700 or 701 in record, make subfield 200f from all fields 700 and/or 701 that have no more than 2 subfields. **Example: ID=19** 702 01**a**Althammer **b**Walter **4**Herausgegeben von 200 0 **f**Herausgegeben von Walter Althammer
						a/R b/R		Entry element Part of name	C	Ignore fields with more than 2 subfields. Convert fields by following rules: - Keep order of conversion according to the fields in original record. - Delete blank from the beginning and from the end of the string. - If there are pair of subfields **ab** in the field, convert subfield **b** before subfield **a** Add blank before subfield **a**. - If there are not pair of subfields **ab**, convert subfield/subfields in order of original record. Add blank before next subfield. - Add comma and blank ", " before next converted field. **Example**: ID =121

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										700 _1 **a**Antoine **b**Serge 701 11 **a**Vilmorin **b**Jean-Baptiste de 701 11 **a**Yana **b**André 200 0_ **f**Serge Antoine, Jean-Baptiste de Vilmorin, André Yana
200/NR		**g/R**		Subsequent statement of responsibility	**702/R**					Data from fields 702 are used for conversion in subfield 200g if there is at least one field 700 or 701 in the record. If there is no field 700 or 701 in record, do not convert fields 702 into 200g. After the conversion into field 200, keep field 702 for following conversions.
						a/R **b/R** **4/R**			A	If no field 702 in the record contains subfield **4,** ignore all fields 702 in that record. Also ignore all fields 702 that contain three or more subfields but without subfield 4. Convert fields by following rules: • Make one subfield 200g from each field 702 with subfield 4 and all following fields 702 without subfield 4. Repeat subfield 200g for every next field 702 with subfield 4 and all following fields 702 without subfield 4. • Keep order of conversion according to the fields in original record. • Add comma and blank ", " before next converted field

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										within one subfield 200g.
										• Convert existing subfields by following order: 4ab **4ba.** Add blank before each added subfield (subfields **ab**). If there is not subfield 4, convert subfield ab; if there is only subfield a or subfield b, convert it.
										Example: ID 240
										702 01 **a**Murra **b**Gj. **4**Përkth. - 702 01 **a**Gjergji **b**B **4**Red. - 702 01 **a**Leka **b**SH.
										200 0_ **g**përkth. Gj. Murra **g**Red. B Gjergji, Sh. Leka
200/NR		**h/R**		Number of part	**200/NR**	**h/R**		Number of part	K	
		i/R		Name of part		**i/NR**		Name of part	K	For adding NSB/NSE at the beginning of subfield 200i, use the definition from the doc. below [Definition for using NSB/NSE](#)
		i/R		Name of part				Name of part	K	
205/NR	--			**EDITION STATEMENT**	**205/NR**			**EDITION STATEMENT**		Once converted data from field 205 is not used any more
		a/NR				**a/NR**		Edition statement	K	Delete "(" and/or ")" from the beginning and from the end of string
210/NR	--			**PUBLICATION, DISTRIBUTION, ETC.**	**210/R**			**PUBLICATION**		• If field 210 is repeated, convert only first field 210. • Other repeated fields 210 convert into repeated subfields 993a. • Add ″_;_″ before data from subfields c and d. • Once converted data from field

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										210 is not used any more
		a/R		Place of publication, distribution, etc.		a/R		Place of publication	C	Parse the subfield according to the semicolon ";" punctuation. (It cannot be at the beginning or at the end of the subfield) Parse the whole data according to the existing punctuation unless there is "etj." or "_etj." after punctuation. In that case covert whole string into subfield 210a Before converting each part of the parsed string into repeated subfield 210a, replace every "T." with "Tiranë". Do not transfer the punctuation ";". The order of the converted subfields must correspond to the order of data in the original record. If the subfield does not contain semicolon ";", replace "T." with "Tiranë" and then convert 210a into 210a.
		c/R		Name of publisher, distributor, etc.		c/R		Name of publisher	C	Parse the subfield according to the semicolon ";" punctuation. (It cannot be at the beginning or at the end of the subfield) Parse the whole data according to the existing punctuation unless there is "etj." or "_etj." after punctuation. In that case covert whole string into subfield 210c Convert each part of the parsed string into repeated subfield ~~200c~~ 210c. Do not transfer the

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										punctuation ";". The order of the converted subfields must correspond to the order of data in the original record. If the subfield do not contain semi colon ";"convert 200c into 200c. For adding NSB/NSE at the beginning of subfield 210c, use the definition from the doc. below Definition for using NSB/NSE
		d/NR		Date of publication, distribution, etc.		d/R		Date of publication	C	Convert only the first subfield 210d Do not convert "." punctuation from the end of the string.
215/R	--			PHYSICAL DESCRIPTION	215/NR			PHYSICAL DESCRIPTION		Once converted data from field 215 is not used any more
		a/NR		Specific material designation		a/R		Specific material designation	K	Convert data from all repeated subfields 215a into nonrepeated subfield 215a. Add blank before data from next subfield 215a.
		a/NR		Specific material designation				Specific material designation		
		c/NR		Other physical details		c/NR		Other physical details	K	
		d/NR		Dimensions		d/NR		Dimensions	C	Do not convert "." punctuation from the end of the string
		e/R		Accompanying material		e/NR		Accompanying material	K	
225/R				SERIES	225/NR			SERIES		Once converted data from field 225 is not used any more.
	1_									
		a/NR		Series title		a/NR		Series title	K	For adding NSB/NSE at the beginning of subfield 225a, use the definition from the doc. below

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
										Definition for using NSB/NSE
		v/R		Volume of designation		v/NR		Volume of designation	K	
		x/R		ISSN of series		x/NR		ISSN of series	K	
300/R				**GENERAL NOTE**	300/R			**GENERAL NOTE**		Convert field 300 into field 300 or 320. Once converted data from field 300 is not used any more
	1_									
		a/NR				a/NR		Text of Note	C	Parse the subfield according to the semicolon ″_;_″ and/or full stop and dach″.-″ punctuation (they cannot be at the beginning or at the end of the subfield)
320/R				**INTERNAL BIBLIOGRAPHIES/INDEXES NOTE**						Parse the whole data according to the existing punctuation.
	1_									Delete blank from the beginning and from the end of string.
		a/NR		**Text of Note**						If parsed string starts with "bibl*" or "Inde*", convert it into repeatable field 320a. Otherwise convert parsed string into repeatable field 300a. The order of the converted fields must be increasing. If the subfield do not contain punctuation but string starts with "bibli*" or "Inde*", convert string into repeatable subfield 320a. Otherwise convert string into repeatable field 300a.
300/R				**GENERAL NOTE**	304/NR			**NOTES PERTAINING TO**		Once converted data from field is

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
								TITLE AND STATEMENT OF RESPONSIBILITY		not used any more
	1_									
		a/NR		Text of Note		a/NR		Text of Note	C	
300/R				**GENERAL NOTE**	305/R			**NOTES PERTAINING TO EDITION AND BIBLIOGRAPHIC HISTORY**		Once converted data from field is not used any more
	1_									
		a/NR		Text of Note		a/NR		Text of Note	C	
300/R				**GENERAL NOTE**	312/NR			**NOTES PERTAINING TO RELATED TITLES**		Once converted data from field is not used any more
	1_									
		a/NR				a/NR		Text of Note	C	
300/R				**GENERAL NOTE**	314/NR			**NOTE PERTAINING TO INTELLECTUAL RESPONSIBILITY**		Once converted data from field is not used any more
	1_									
		a/NR		Text of note		a/NR		Text of note	C	
300/R				**GENERAL NOTE**	327/NR			**CONTENTS NOTE**		Once converted data from field is not used any more
	1_									
		a/NR		Text of note		a/NR		Text of note	C	
320/R				**INTERNAL BIBLIOGRAPHIES/INDEXES NOTE**	320/NR			Text of note		Once converted data from field is not used any more
	1_									
		a/NR		Text of note		a/NR		Text of note	K	
330/R	--			**ABSTRACT**	330/R			**ABSTRACT**		Once converted data from field is not used any more
		a/NR		Text of note		a/NR		Text of note	K	
		z/NR		Language						
			alb						A	Add code **alb** to all records

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
600/R				**PERSONAL NAME USED AS SUBJECT**	600/R			**PERSONAL NAME USED AS SUBJECT**		Once converted data from field 600 is not used any more.
	10									
		a/NR		Entry element		**a/NR**		Entry element	K	The first letter in the string always converts into capital letter if it is not. Delete punctuation from the end of string.
		b/NR		Part of name (not entry element)		**b/NR**		Part of the name	K	The first letter in the string always converts into capital letter if it is not.
607/R				**GEOGRAPHICAL NAME USED AS SUBJECT**	607/R			**GEOGRAPHICAL NAME AS USED AS SUBJECT**		Once converted data from field 607 is not used any more.
	1_									
		a/NR		Entry element		**a/NR**		Entry element	K	
610/R				**UNCONTROLLED SUBJECT TERMS**	610/R			**UNCONTROLLED SUBJECT TERM**		Once converted data from field 610 is not used any more. Convert repeated field 610a into repeated subfield 610a
	0_									
		a/R		Subject term		**a/NR**		Subject term	K	Do not convert "." punctuation from the end of the string
		z/NR		Language						
			alb						A	Add code **alb** to all records
675/R	**--**			**UNIVERSAL DECIMAL CLASSIFICATION (UDC)**	675/R			**UNIVERSAL DECIMAL CLASSIFICATION (UDC)**		Once converted data from field 675 is not used any more.
		a/NR		Number		**a/NR**		Number	K	Delete all spaces from the string.
		c/NR		UDC access*					A	
			fik							
700/NR				**PERSONAL NAME – PRIMARY INTELLECTUAL**	700/NR			**PERSONAL NAME – PRIMARY**		Once converted data from field 700 is not used any more.

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
				RESPONSIBILITY				INTELLECTUAL RESPONSIBILITY		
	_1									If there is subfield 700b
	_0									If there is not subfield 700b
		a/NR b/NR		Entry element		a/R		Entry element	K	If the subfield 700a is repeated, convert repeated subfield 700a to the subfield 700b
		b/NR		Part of name (not entry element)		b/NR		Part of name (not entry element)	K	
		4/R		Relator code					A	
			070							Add the code 070 as default to all records.
701/R				PERSONAL NAME– ALTERNATIVE INTELLECTUAL RESPONSIBILITY	701/R			PERSONAL NAME- ALTERNATIVE INT. RESPONSIBILITY		Once converted data from field 701 is not used any more.
	11 01									If there is subfield 701b, first field 701 has indicators [11], all other repeated fields [01]
	10 00									If there is not subfield 701b, first field 701 has indicators [10], all other repeated fields [00]
		a/NR		Entry element		a/R		Entry element	K	If the subfield 701a is repeated convert repeated subfield 701a to the subfield 701b Delete ˝_etj.˝ if exists in the subfield 701a
		b/NR		Part of name (not entry element)		b/NR		Part of name (not entry element)	K	Delete ˝_etj.˝ if exists in the subfield 701b
		4/R		Relator code					A	
			070							Add the code 070 as default to all records.
702/R				PERSONAL NAME –	702/R			PERSONAL NAME- SECONDARY INT.		Once converted data from field 702 is not used any more.

COMARC/B					WINISIS FORMAT/B				Status	NOTES
Field	Ind. value	Subfield	Code	Meaning of subfield	Field	Subfield	Code	Meaning of subfield		
				SECONDARY INTELLECTUAL RESPONSIBILITY				RESPONSIBILITY		
	01									If there is subfield 702b
	00									If there is not subfield 702b
		a/NR		Entry element		a/NR		Entry element	K	Delete ″_etj.″ if exists in the subfield 702a
		b/NR		Part of name (not entry element)		b/NR		Part of name	K	Delete ″_etj.″ if exists in the subfield 702a
		4/R		Relator code		4/R			C	If there are following data in subfield 7024 add adequate data as below: bashkepun* 070 bashkëpun* 070 perkth* 730 përkth* 730 pergat* 220 përgat* 220 shqiper* 730 shqipër* 730 edit* 340 ured* 340 redak* 340 red. 340 ed. 340 traduz* 730 transla* 730
710/NR				CORPORATE BODY NAME – PRIMARY INTELLECTUAL RESPONSIBILITY	710/NR			CORPORATE BODY NAME-PRIM. INTELLECTUAL RESPONSIBILITY		Once converted data from field 710 is not used any more.
	02									
		a/NR		Entry element		a/NR		Entry element	K	
		e/R		Location of meeting		e/NR		Location of meeting	K	
		f/NR		Date of meeting		f/NR		Date of meeting	K	

Legende:

A – Add
C – Change
K – Keep

APPENDIX B: Database statistics

Tag	S	Freq	MinR	MaxR	MinT	MaxT	Avrg1	Avrg2	Uniq	UniqA	IdR
000		6839	1	1			1.000	1.000	6839	1.000	0
	a	6839	1	1	1	1	1.000	1.000	6839	1.000	0
	b	6839	1	1	1	1	1.000	1.000	6839	1.000	0
	c	6839	1	1	1	1	1.000	1.000	6839	1.000	0
	d	6838	1	1	1	1	0.999	1.000	6838	0.999	1
	e	6838	1	1	1	1	0.999	1.000	6838	0.999	1
	f	6838	1	1	1	1	0.999	1.000	6838	0.999	1
	g	6838	1	1	1	1	0.999	1.000	6838	0.999	1
006		11242	1	18			1.643	1.651	6809	0.995	4106
	a	11242	1	18	1	1	1.643	1.651	6809	0.995	4106
010		4751	1	4			0.694	1.052	4512	0.659	3779
	a	3829	1	4	1	2	0.559	1.063	3600	0.526	5143
	b	22	1	2	1	1	0.003	1.375	16	0.002	65
	d	3044	1	4	1	1	0.445	1.005	3028	0.442	3779
021		2	1	1			0.000	1.000	2	0.000	1459
	a	2	1	1	1	1	0.000	1.000	2	0.000	1459
	b	2	1	1	1	1	0.000	1.000	2	0.000	1459
022		2	1	1			0.000	1.000	2	0.000	1490
	b	2	1	1	1	1	0.000	1.000	2	0.000	1490
100		50	1	1			0.007	1.000	50	0.007	1458
	a	50	1	1	1	1	0.007	1.000	50	0.007	1458
101		3127	1	4			0.457	1.002	3119	0.456	1486
	a	3131	1	4	1	2	0.457	1.003	3119	0.456	1486
	c	2875	1	1	1	1	0.420	1.000	2875	0.420	1
	f	2	1	1	1	1	0.000	1.000	2	0.000	1463
102		51	1	2			0.007	1.020	50	0.007	1489
	a	51	1	2	1	1	0.007	1.020	50	0.007	1489
105		34	1	1			0.004	1.000	34	0.004	1459
	a	34	1	1	1	1	0.004	1.000	34	0.004	1459
106		35	1	1			0.005	1.000	35	0.005	1459
	a	35	1	1	1	1	0.005	1.000	35	0.005	1459
200		6838	1	1			0.999	1.000	6838	0.999	1
	a	6842	1	2	1	2	1.000	1.000	6838	0.999	2419
	e	3376	1	3	1	3	0.493	1.005	3358	0.491	4698
	h	1187	1	2	1	2	0.173	1.000	1186	0.173	5726
	i	550	1	1	1	1	0.080	1.000	550	0.080	85
205		396	1	1			0.057	1.000	396	0.057	29
	a	396	1	1	1	1	0.057	1.000	396	0.057	29
210		6924	1	3			1.012	1.014	6824	0.997	2941
	a	6870	1	3	1	2	1.004	1.009	6804	0.994	3544

```
Statistics database of the fields/subfields before conversion                          2
--------------------------------------------------------------------------------
Tag S     Freq  MinR  MaxR  MinT  MaxT  Avrg1  Avrg2   Uniq  UniqA        IdR
--------------------------------------------------------------------------------
    c     5841    1     2     1     2   0.854  1.012   5768  0.843        830
    d     6688    1     3     1     2   0.977  1.002   6673  0.975       3544

215       6238    1     1               0.912  1.000   6238  0.912          1
    a     6256    1     2     1     2   0.914  1.003   6234  0.911        444
    c     1504    1     1     1     1   0.219  1.000   1504  0.219          1
    d     5501    1     1     1     1   0.804  1.000   5501  0.804          1
    e       25    1     1     1     1   0.003  1.000     25  0.003       1503
    A        1    1     1     1     1   0.000  1.000      1  0.000       4514

225       2630    1     1               0.384  1.000   2630  0.384          1
    a     2579    1     1     1     1   0.377  1.000   2579  0.377          1
    v      366    1     1     1     1   0.053  1.000    366  0.053        789
    x      127    1     1     1     1   0.018  1.000    127  0.018        163

300       4490    1     4               0.656  1.331   3371  0.492       1018
    a     4490    1     4     1     1   0.656  1.331   3371  0.492       1018

301       6823    1     1               0.997  1.000   6823  0.997          1
    a     6823    1     1     1     1   0.997  1.000   6823  0.997          1

302       6804    1     1               0.994  1.000   6804  0.994          1
    a     6804    1     1     1     1   0.994  1.000   6804  0.994          1

304          1    1     1               0.000  1.000      1  0.000       1458
    a        1    1     1     1     1   0.000  1.000      1  0.000       1458

305          3    1     2               0.000  1.500      2  0.000       1486
    a        3    1     2     1     1   0.000  1.500      2  0.000       1486

312          1    1     1               0.000  1.000      1  0.000       1493
    a        1    1     1     1     1   0.000  1.000      1  0.000       1493

314          1    1     1               0.000  1.000      1  0.000       2553
    a        1    1     1     1     1   0.000  1.000      1  0.000       2553

320          1    1     1               0.000  1.000      1  0.000       1497
    a        1    1     1     1     1   0.000  1.000      1  0.000       1497

327          1    1     1               0.000  1.000      1  0.000       1489
    a        1    1     1     1     1   0.000  1.000      1  0.000       1489

330          5    1     1               0.000  1.000      5  0.000         83
    a        5    1     1     1     1   0.000  1.000      5  0.000         83

410          4    1     1               0.000  1.000      4  0.000       1463
    a        4    1     1     1     1   0.000  1.000      4  0.000       1463
    1        4    1     1     1     1   0.000  1.000      4  0.000       1463

421          1    1     1               0.000  1.000      1  0.000       1473
    a        1    1     1     1     1   0.000  1.000      1  0.000       1473
    1        1    1     1     1     1   0.000  1.000      1  0.000       1473
```

```
Statistics database of the fields/subfields before conversion                    3
--------------------------------------------------------------------------------
Tag S      Freq  MinR  MaxR  MinT  MaxT  Avrg1 Avrg2    Uniq  UniqA        IdR
--------------------------------------------------------------------------------
422           1     1     1               0.000 1.000       1  0.000       1469
     a        1     1     1     1     1   0.000 1.000       1  0.000       1469
     1        1     1     1     1     1   0.000 1.000       1  0.000       1469

510           5     1     2               0.000 1.250       4  0.000       1496
     a        5     1     2     1     1   0.000 1.250       4  0.000       1496
     z        5     1     2     1     1   0.000 1.250       4  0.000       1496

512           2     1     1               0.000 1.000       2  0.000       1479
     a        2     1     1     1     1   0.000 1.000       2  0.000       1479

600         221     1     5               0.032 1.104     200  0.029        736
     a      221     1     5     1     1   0.032 1.104     200  0.029        736
     b      168     1     3     1     1   0.024 1.076     156  0.022       2609

601           1     1     1               0.000 1.000       1  0.000       1466
     a        1     1     1     1     1   0.000 1.000       1  0.000       1466
     e        1     1     1     1     1   0.000 1.000       1  0.000       1466
     f        1     1     1     1     1   0.000 1.000       1  0.000       1466
     x        1     1     1     1     1   0.000 1.000       1  0.000       1466
     2        1     1     1     1     1   0.000 1.000       1  0.000       1466

606          14     1     3               0.002 1.750       8  0.001       1484
     a        8     1     1     1     1   0.001 1.000       8  0.001       1466
     x       14     1     3     1     1   0.002 1.750       8  0.001       1484
     y        2     1     1     1     1   0.000 1.000       2  0.000       1488
     2        4     1     1     1     1   0.000 1.000       4  0.000       1466

607         930     1     6               0.135 1.336     696  0.101       3724
     a      930     1     6     1     1   0.135 1.336     696  0.101       3724

610       12769     1    30               1.867 1.871    6822  0.997       3562
     a    12769     1    30     1     1   1.867 1.871    6822  0.997       3562

675       11265     1     7               1.647 1.651    6821  0.997       3210
     a    11278     1     7     1     2   1.649 1.653    6821  0.997       3210
     v     1364     1     5     1     2   0.199 1.705     800  0.116       1466
     z     1268     1     5     1     2   0.185 1.720     737  0.107       1466
     A        2     1     1     1     1   0.000 1.000       2  0.000       5735
     V        3     1     1     1     1   0.000 1.000       3  0.000       1574
     Z        7     1     2     1     1   0.001 1.399       5  0.000       5735

700        4699     1     1               0.687 1.000    4699  0.687          6
     a     4702     1     2     1     2   0.687 1.000    4698  0.686        123
     b     4656     1     1     1     1   0.680 1.000    4656  0.680          6

701        1003     1    10               0.146 1.597     628  0.091       3533
     a     1006     1    10     1     2   0.147 1.601     628  0.091       3533
     b      987     1    10     1     2   0.144 1.591     620  0.090       3533

702        4024     1    15               0.588 1.371    2935  0.429       5776
     a     4022     1    15     1     2   0.588 1.370    2934  0.429       5776
```

```
Statistics database of the fields/subfields before conversion                    4
--------------------------------------------------------------------------------
Tag S      Freq  MinR  MaxR  MinT  MaxT  Avrg1  Avrg2    Uniq  UniqA       IdR
--------------------------------------------------------------------------------
    b      3998     1    15     1     1  0.584  1.366    2926  0.427      5776
    4      3327     1     5     1     1  0.486  1.173    2835  0.414      2951

710           6     1     1                0.000  1.000       5  0.000      4593
    a         6     1     1     1     1  0.000  1.000       5  0.000      4593
    e         1     1     1     1     1  0.000  1.000       1  0.000      5885
    f         1     1     1     1     1  0.000  1.000       1  0.000      5885

712          18     1     2                0.002  1.058      17  0.002      1466
    a        18     1     2     1     1  0.002  1.058      17  0.002      1466
    b         6     1     1     1     1  0.000  1.000       6  0.000      1475
    4        18     1     2     1     1  0.002  1.058      17  0.002      1466

801          50     1     1                0.007  1.000      50  0.007      1458
    a        50     1     1     1     1  0.007  1.000      50  0.007      1458
    b        50     1     1     1     1  0.007  1.000      50  0.007      1458

922          51     1     1                0.007  1.000      51  0.007      1458
    a        51     1     1     1     1  0.007  1.000      51  0.007      1458


================================================================================
Legend:

Tag - Field code (three-digit code)
S - Subfield code (one-digit code)
Freq - Occurrences of fields/subfields in the processed records
MinR - The minimum number of occurrences of fields/subfields in a single record
MaxR - The maximum number of occurrences of fields/subfields in a single record
MinT - The minimum number of occurrences of subfields within a field in the record
MaxT - The maximum number of occurrences of subfields within a field in the record
Uniq - The number of records in which a field/subfield occurs
IdR - The record in which a field/subfield is repeated several times
```

```
Statistics database of the fields/subfields after conversion                    1
--------------------------------------------------------------------------------
Tag S      Freq  MinR  MaxR  MinT  MaxT  Avrg1  Avrg2    Uniq  UniqA        IdR
--------------------------------------------------------------------------------
000        6839     1     1               1.000  1.000    6839  1.000          0
    a      6839     1     1     1     1   1.000  1.000    6839  1.000          0
    b      6839     1     1     1     1   1.000  1.000    6839  1.000          0
    c      6839     1     1     1     1   1.000  1.000    6839  1.000          0
    d      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    e      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    f      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    g      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    t      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    u      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    v      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520

001        6838     1     1               0.999  1.000    6838  0.999   1536001520
    a      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    b      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    c      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    d      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    e      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    7      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520

010        4745     1     4               0.693  1.053    4506  0.658   1536968688
    a      3826     1     4     1     2   0.559  1.063    3599  0.526   1537317872
    b       303     1     4     1     1   0.044  1.578     192  0.028   1537317872
    d      3021     1     4     1     1   0.441  1.005    3005  0.439   1536968688

100        6838     1     1               0.999  1.000    6838  0.999   1536001520
    b      6450     1     1     1     1   0.943  1.000    6450  0.943   1536001520
    c      6450     1     1     1     1   0.943  1.000    6450  0.943   1536001520
    d         5     1     1     1     1   0.000  1.000       5  0.000   1536115184
    e      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    g      3119     1     1     1     1   0.456  1.000    3119  0.456   1536001520
    h      6838     1     1     1     1   0.999  1.000    6838  0.999   1536001520
    i      3119     1     1     1     1   0.456  1.000    3119  0.456   1536001520
    l      3119     1     1     1     1   0.456  1.000    3119  0.456   1536001520

101        3101     1     1               0.453  1.000    3101  0.453   1536001520
    a      3103     1     4     1     4   0.453  1.003    3092  0.452   1536381680
    c       607     1     1     1     1   0.088  1.000     607  0.088   1536198384
    f         2     1     1     1     1   0.000  1.000       2  0.000   1536375792

102        6034     1     1               0.994  1.000    6034  0.994   1536001520
    a      6034     1     2     1     2   1.000  1.005    6034  0.994   1536097776

105        2362     1     1               0.345  1.000    2362  0.345   1536001520
    a      1504     1     1     1     1   0.219  1.000    1504  0.219   1536001520
    e      1276     1     1     1     1   0.186  1.000    1276  0.186   1536002288

106          35     1     1               0.005  1.000      35  0.005   1536374768
    a        35     1     1     1     1   0.005  1.000      35  0.005   1536374768

200        6838     1     1               0.999  1.000    6838  0.999   1536001520
    a      6842     1     2     1     2   1.000  1.000    6838  0.999   1536620528
```

```
Statistics database of the fields/subfields after conversion                     2
```

Tag	S	Freq	MinR	MaxR	MinT	MaxT	Avrg1	Avrg2	Uniq	UniqA	IdR
	d	233	1	6	1	6	0.034	1.059	220	0.032	1537649648
	e	3963	1	11	1	11	0.579	1.162	3408	0.498	1537644528
	f	6248	1	1	1	1	0.913	1.000	6248	0.913	1536002288
	g	1871	1	4	1	4	0.273	1.262	1482	0.216	1536756720
	h	1187	1	2	1	2	0.173	1.000	1186	0.173	1537467120
	i	550	1	1	1	1	0.080	1.000	550	0.080	1536023024
205		396	1	1			0.057	1.000	396	0.057	1536008688
	a	396	1	1	1	1	0.057	1.000	396	0.057	1536008688
210		6824	1	1			0.997	1.000	6824	0.997	1536001520
	a	7042	1	3	1	3	1.029	1.034	6804	0.994	1536156400
	c	5823	1	3	1	3	0.851	1.013	5745	0.840	1536218608
	d	6640	1	1	1	1	0.970	1.000	6640	0.970	1536001520
215		6237	1	1			0.911	1.000	6237	0.911	1536001520
	a	6232	1	1	1	1	0.911	1.000	6232	0.911	1536001520
	c	1504	1	1	1	1	0.219	1.000	1504	0.219	1536001520
	d	5501	1	1	1	1	0.804	1.000	5501	0.804	1536001520
	e	25	1	1	1	1	0.003	1.000	25	0.003	1536386032
225		2629	1	1			0.384	1.000	2629	0.384	1536001520
	a	2578	1	1	1	1	0.376	1.000	2578	0.376	1536001520
	v	366	1	1	1	1	0.053	1.000	366	0.053	1536203248
	x	127	1	1	1	1	0.018	1.000	127	0.018	1536042992
300		1003	1	4			0.146	1.108	905	0.132	1536329456
	a	1003	1	4	1	1	0.146	1.108	905	0.132	1536329456
320		3495	1	3			0.511	1.275	2740	0.400	1536384496
	a	3495	1	3	1	1	0.511	1.275	2740	0.400	1536384496
330		5	1	1			0.000	1.000	5	0.000	1536022512
	a	5	1	1	1	1	0.000	1.000	5	0.000	1536022512
	z	5	1	1	1	1	0.000	1.000	5	0.000	1536022512
600		221	1	5			0.032	1.104	200	0.029	1536189680
	a	221	1	5	1	1	0.032	1.104	200	0.029	1536189680
	b	168	1	3	1	1	0.024	1.076	156	0.022	1536669168
607		930	1	6			0.135	1.336	696	0.101	1536954608
	a	930	1	6	1	1	0.135	1.336	696	0.101	1536954608
610		12769	1	30			1.867	1.871	6822	0.997	1536913136
	a	12768	1	30	1	1	1.866	1.871	6822	0.997	1536913136
	z	12769	1	30	1	1	1.867	1.871	6822	0.997	1536913136
675		11264	1	7			1.647	1.651	6821	0.997	1536823024
	a	11278	1	7	1	2	1.649	1.653	6821	0.997	1536823024
	c	11280	1	7	1	2	1.649	1.653	6821	0.997	1536823024
700		4699	1	1			0.687	1.000	4698	0.687	1536002800

```
Statistics database of the fields/subfields after conversion                     3
--------------------------------------------------------------------------------
Tag S      Freq  MinR  MaxR  MinT  MaxT  Avrg1 Avrg2   Uniq  UniqA        IdR
--------------------------------------------------------------------------------
    a      4698     1     1     1     1  0.686 1.000   4698  0.686 1536002800
    b      4660     1     1     1     1  0.681 1.000   4656  0.681 1536002800
    4      4699     1     1     1     1  0.687 1.000   4699  0.687 1536002800

701        1003     1    10                 0.146 1.597    628  0.091 1536905712
    a       998     1    10     1     1  0.145 1.599    628  0.091 1536905712
    b       991     1    10     1     2  0.144 1.593    620  0.090 1536905712
    4      1003     1    10     1     1  0.146 1.597    628  0.091 1536905712

702        4019     1    15                 0.587 1.369   2935  0.429 1537479920
    a      4017     1    15     1     2  0.587 1.369   2934  0.429 1537479920
    b      3998     1    15     1     1  0.584 1.366   2926  0.427 1537479920
    4      2744     1     5     1     2  0.401 1.173   2338  0.341 1536756720

710           5     1     1                 0.000 1.000      5  0.000 1537177072
    a          5     1     1     1     1  0.000 1.000      5  0.000 1537177072
    e          1     1     1     1     1  0.000 1.000      1  0.000 1537507824
    f          1     1     1     1     1  0.000 1.000      1  0.000 1537507824

993         100     1     2                 0.014 1.052     95  0.013 1536754160
    a        101     1     2     1     2  0.014 1.063     95  0.013 1536381680

996       11239     1    18                 1.643 1.651   6807  0.995 1537052400
    d     11239     1    18     1     1  1.643 1.651   6807  0.995 1537052400
    f     11175     1    18     1     1  1.634 1.655   6750  0.986 1537052400
    o     11239     1    18     1     1  1.643 1.651   6807  0.995 1537052400
    p     11239     1    18     1     1  1.643 1.651   6807  0.995 1537052400
    r        48     1     2     1     1  0.007 1.066     45  0.006 1537027568
    t     11239     1    18     1     1  1.643 1.651   6807  0.995 1537052400
    w     11239     1    18     1     1  1.643 1.651   6807  0.995 1537052400
    6     11239     1    18     1     1  1.643 1.651   6807  0.995 1537052400


================================================================================

Legend:

Tag - Field code (three-digit code)
S - Subfield code (one-digit code)
Freq - Occurrences of fields/subfields in the processed records
MinR - The minimum number of occurrences of fields/subfields in a single record
MaxR - The maximum number of occurrences of fields/subfields in a single record
MinT - The minimum number of occurrences of subfields within a field in the record
MaxT - The maximum number of occurrences of subfields within a field in the record
Uniq - The number of records in which a field/subfield occurs
IdR - The record in which a field/subfield is repeated several times
```

APPENDIX C: Country publication code

Place of publication	Country code
Acron	usa
Albinform	alb
Allea	nld
Amsterdam	nld
Anacapri	ita
Ankara	tur
Arhus	dnk
Aris	fra
Arizona	usa
Arles	fra
Arles	fra
Athena	grc
Athens	grc
Athina	grc
Athinë	grc
Athones	grc
Athőnes	grc
Auckland	nzl
Baku	aze
Baltimore	usa
Barcalona	esp
Barcelona	esp
Barcelone	esp
Bari	ita
Basel	che
Basingstoke	gbr
Bayreuth	deu
Beijing	chn
Belgrad	srb
Belgrade	srb
Belmont	usa
Beograd	srb
Berat	alb
Bergamo	ita
Berkeley	usa
Berkley	usa
Berlin	deu
Berling	deu

Place of publication	Country code
Bern	che
Berne	che
Betat	alb
Bilthoven	nld
Blijde	bel
Bloomington	usa
Bochu	deu
Bochum	deu
Bologna	ita
Bonn	deu
Bordeaux	fra
Boston	usa
Bratislava	svk
Brescia	ita
Bristol	gbr
Brixelles	bel
Brno	cze
Bruxelles	bel
Bucarest	rou
Bucharest	rou
Bucureshti	rou
Bucuresti	rou
Budapest	hun
Bukuresht	rou
Cacak	srb
Cairo	egy
Califonia	usa
California	usa
Cambridge	gbr
Castrovillari	ita
Catanzaro	ita
Cedex	fra
Celjabinsk	rus
Cédex	fra
Champaign	usa
Chatham	usa
Chicago	usa
Chichester	gbr
Clermont	usa
Cleveland	usa
Colombus	usa
Colorado	usa

Place of publication	Country code
Columbus	usa
Connecticut	usa
Copenhagen	dnk
Cortona	ita
Cosenza	ita
Craiova	rou
Cremona	ita
Çaçak	srb
Çeljabinsk	rus
Dallas	usa
Darmstadt	deu
Deenver	usa
Denmark	dnk
Denver	usa
Donetsk	ukr
Dordrecht	nld
Dubrovnik	hrv
Dumont	usa
Durham	usa
Durrës	alb
Edinburgh	gbr
Elbasan	alb
Emmaus	gbr
England	gbr
Englewood	usa
Erlangen	deu
Faenza	ita
Fier	alb
Fiesole	ita
Firence	ita
Firenze	ita
Flers	fra
Florida	usa
Foggia	ita
Frankfurt	deu
Geneva	che
Geneve	che
Genéve	che
Genova	ita
Gent-leuven	bel
Germany	deu
Gjenevë	che

Place of publication	Country code
Gjirokastër	alb
Glenview	usa
Gorizia	ita
Gotingen	ger
Gottingen	ger
Götingen	ger
Göttingen	ger
Graz	aus
Grenoble	fra
Gutersloh	deu
Gütersloh	deu
Hague	nld
Hamburg	deu
Hamilton	can
Harjkov	rus
Helsinki	fin
Homewood	usa
Huddersfield	gbr
Hull	can
Huntsville	usa
Ilford	gbr
Illinois	usa
Illinos	usa
Indianapolis	usa
Innsbrick	aut
Innsbruck	aut
Insbruck	aut
Iowa	usa
Ipswich	gbr
Istanbul	tur
Italia	ita
Janinë	grc
Jefferson	usa
Jersey	usa
Karachi	pak
Kiel	deu
Kiev	ukr
Kishinev	mda
Kjrdzhali	bgr
Koln	deu
Konigheim	deu
Korçë	alb

Place of publication	Country code
Korfuz	grc
Köln	deu
Königheim	deu
Krakow	pol
Krakowie	pol
Kraskow	pol
Kwell	gbr
La garenne-colombus	fra
La salle	usa
Lanciano	ita
Lattes	fra
Lausanne	che
Lavoisier	fra
Lecce	ita
Lefkosa	cyp
Leipzig	deu
Leningrad	rus
Leninigrad	rus
Leuven	bel
Lewiston	usa
Lexington	usa
Lezhë	alb
Liebefels-berne	che
Liege	bel
Liége	bel
Ljubjana	svn
Ljubljana	svn
Ljubljna	svn
Locarno	che
Lodz	pol
Londër	gbr
Londo	gbr
London	gbr
Los angeles	usa
Lódz	pol
Lubjana	svn
Luxembourg	lux
Lyon	fra
M.	rus
Maastricht	nld
Macon	usa
Madrid	esp

Place of publication	Country code
Malmo	swe
Mannheim	deu
Manz	deu
Maplewood	usa
Marburg	deu
Marseille	fra
Maryland	usa
Massachusetts	usa
Masson	fra
Miami	usa
Michigan	usa
Milano	ita
Milot	alb
Minneapolis	usa
Minsk	blr
Montreal	ca
Montréal	ca
Morgantown	usa
Moscou	rus
Moscow	rus
Moskav	rus
Moskë	rus
Moskow	rus
Moskva	rus
Moskva6cmislj	rus
Mullet	alb
Munchen	deu
München	deu
Mynih	deu
Nagold	deu
Nairobi	ken
Nantes	fra
Naples	ita
Napoli	ita
Nederland	nld
Neuchâtel	che
New brunswick	ca
New haven	gbr
New heaven	gbr
New jersey	usa
New london	usa
New providence	usa

Place of publication	Country code
New providence new jersey	usa
New york	usa
Në	alb
Ngrad	rus
Nijimegen	nld
Northumberland	gbr
Noth-holland-amsterdam	nld
Novara	ita
Novi sad	srb
Novosi-birsk	rus
Novosibirsk	rus
Novosibirski	rus
Num press	usa
Ohio	usa
Oklahoma	usa
Oksford	gbr
Olten	che
Ontario	ca
Opava	cze
Oslo	nor
Ottawa	ca
Oxfor; new york	gbr
Oxford	gbr
Padova	ita
Pakistan	pak
Palermo	ita
Palo del colle ba itali	ita
Paric	fra
Paris	fra
Paris6cmasson	fra
Patos	alb
Pekino	chn
Pennsylvania	usa
Pfaffenweiler	deu
Philadelphia	usa
Pisa	ita
Pittsburgh	usa
Plouzane france	fra
Plouzané france	fra
Podgorica	mne
Pogradec	alb
Potomac	usa

Place of publication	Country code
Potsdam	deu
Povidenc, rhode island	usa
Praha	cze
Princeton	usa
Puiseaux	fra
Pvd.	alb
Pyongyang	kor
Quebec	ca
Qum	irn
Rancho cordova,ca.	ca
Ravenna	ita
Rd	gbr
Regensburg	deu
Remiremont	fra
Rende	ita
Rende cosenza	ita
Rennes	fra
Richardson	usa
Riga	lva
Rimini	ita
Rio de janeiro	bra
Roma	ita
Romania	ita
Romë	ita
Rotterdam	nld
Rouen	fra
Salzburg;wien	aut
San francisco	usa
Sankt-peterburg	rus
Sarajevo	bih
Saveria mannelli catanzaro	ita
Seattle	usa
Shkodër	alb
Shkup	mkd
Shtutgart	deu
Shtuttgart	deu
Siena	fra
Singapor	sgp
Singapore	sgp
Skopje	mkd
Sofia	bgr

Place of publication	Country code
Sofija	bgr
Sopfia	bgr
Southampton	gbr
Soveria mannelli catanzaro	ita
Split	hrv
Spoleto	ita
Springfield	usa
St.paul	usa
Standford	usa
Stanford	usa
Stockholm	swe
Strasbourg	fra
Strasbourg	fra
Stroudsbourg	fra
Stttgarrt	deu
Stuttgard	deu
Stuttgart	deu
Suisse	che
Sunderland	gbr
Sverdlovsk	rus
Switzerland	che
Syracuse	usa
T	alb
T.	alb
Taranto	ita
Tashkent	uzb
Tbilisi	geo
Tetove	mkd
Tetovë	mkd
Texas	usa
Tha hague	nld
The hage	nld
The hague	nld
The university of chicago	usa
Thessalonike	grc
Thessalonikes	grc
Thessaloniki	grc
Tirana	alb
Tirane	alb
Titanë	alb
Titograd	mne

Place of publication	Country code
Tokyo	jpn
Torina	ca
Torino	ita
Toronto	ca
Torun	pol
Totonto-buffalo	ca
Toulouse	fra
Tours	fra
Trelleborg	swe
Tubingen	deu
Tuscon	usa
Tübingen	deu
Ulqin	mne
Uppsala	swe
Urbana	usa
Urbino	ita
Usa	usa
Varshava	pol
Vatican	ita
Vatican city	ita
Ve	ita
Venezia	ita
Venosa	ita
Verona	ita
Vienna	aut
Villeneuve d'ascq	fra
Virginia	usa
Warsau	pol
Warsaw	pol
Warszaw	pol
Warszawa	pol
Washington	usa
Weinheim	deu
Weinhein	deu
Westport	usa
Wien	aut
Wiesbaden	deu
Wilmington	usa
Wisconsin	usa
Wroclaw	pol
Wydawnictwo	pol
York	gbr

Place of publication	Country code
Zadar	hrv
Zagreb	hrv
Zenica	bih
Zurich	che
Zürich	che

APPENDIX D: Mapping table of language code - WINISIS to COMARC/B

WINISIS Language	COMARC/B Code	Script	Modification code	Transliteration code
aangleze	eng	ba	0	y
abgleze	eng	ba	0	y
ahqipe	alb	ba	0	y
alb	alb	ba	0	y
angleze	eng	ba	0	y
anglisht	eng	ba	0	y
arbereshe	alb	ba	0	y
boshnjake	bos	ca	1	b
bullagre	bul	ca	1	b
bullgare	bul	ca	1	b
ceke	cze	ba	0	y
drenge	fre	ba	0	y
eng	eng	ba	0	y
frcnge	fre	ba	0	y
fre	fre	ba	0	y
frege	fre	ba	0	y
frenge	fre	ba	0	y
frengjisht	fre	ba	0	y
frëge	fre	ba	0	y
frënge	fre	ba	0	y
frëngjisht	fre	ba	0	y
ger	ger	ba	0	y
gjermane	ger	ba	0	y
gjuha	alb	ba	0	y
greke	gre	ga	1	b
hungareze	hun	ba	0	y
ipe	alb	ba	0	y
ita	ita	ba	0	y
itali	ita	ba	0	y
italiana	ita	ba	0	y
italiane	ita	ba	0	y
kroate	cro	ca	1	b
maqedon	mac	cc	1	b
maqedonase	mac	cc	1	b
maqedone	mac	cc	1	b
mul	mul	ba	0	y
ng	eng	ba	0	y
nge	eng	ba	0	y
pa	eng	ba	0	y
pe	alb	ba	0	y
polake	pol	ba	0	y

WINISIS Language	COMARC/B Code	Script	Modification code	Transliteration code
polonisht	pol	ba	0	y
por	por	ba	0	y
rumune	rum	ba	0	y
ruse	rus	ca	1	b
serbe	scc	cb	1	b
serbo-kroate	scc, scr	ca	1	b
serbokroate	scc, scr	ca	1	b
serbokroatisht	scc, scr	ca	1	b
shaipe	alb	ba	0	y
shipe	alb	ba	0	y
shqip	alb	ba	0	y
shqipe	alb	ba	0	y
shqipe italiane	alb	ba	0	y
shqipee	alb	ba	0	y
shqipr	alb	ba	0	y
shze	alb	ba	0	y
spa	spa	ba	0	y
spanjolle	spa	ba	0	y
turke	tur	ba	0	y

APPENDIX E: Fields not converted

Field	Subfield	Meaning of fields and subfields	Note
006/		**Copy number**	This field doesn't exist in COMARC/B. It will be converted to COMARC/H
	a/NR		
O21/R		**Legal deposit number**	Field exists only in 2 records and contains irrelevant data
	a/NR	Country code	
	b/NR	Number	
022/R		**Goverment publication number**	Field exists only in 2 records and contains irrelevant data
	b/NR	Number	
100/NR		**General processing data**	Field exists only in 50 records and contains data that will be by default in all records in 100b/c/d/e/h/l
	a/NR	General processing data	
102/NR		**Country of publication or production**	Field exists only in 50 records and contains irrelevant data. This field will be defined for all records according to table (country_publ_code)
	a/NR	Country	
301/R		**Note pertaining to identification number**	In Winisis this field is used for shelf mark number. It will be converted to COMARC/H
	a/NR	Text of note	
302/R		**Location**	In Winisis this field is used for location. It will be converted to COMARC/H
	a/NR	Text of Note	
410/NR		**Series**	Field exists only in 4 record with unknown and irrelevant data
	a/NR	Title proper/Key title	
	1		
421/NR		**Supplement**	Field exists only in 1 record with unknown and irrelevant data
	a/NR		
	1		
422/R		**Parent of supplement**	Field exists only in 1 record with unknown and irrelevant data
	a/NR	Title proper	
510/R		**Paralel title proper**	Field exists only in 4 record with irregular data serves as an example
	a/NR	Paralel title	
512/R		**Cover title**	Field exists only in 2 record with

Field	Subfield	Meaning of fields and subfields	Note
			irregular data serves as an example
	a/NR	Cover title	
601/R		**Corporate body name used as subject**	Field exists only in 1 record with unknown and irrelevant data
	a/Nr	Entry element	
	e/R	Location of meeting	
	x/R	Topical subdivision	
	2/NR	System code	
606/R		**Topical name used as subject**	Field exists only in 14 record with unknown and irrelevant data
	a/R	Entry element	
	x/R	Topical subdivision	
	y/R	Geographical subdivision	
	2/NR	System code	
675/R		**Universal decimal classification (UDC)**	
	v/NR	Edition	
	z/NR	Language of edition	
712/NR		**Corporate body name – secondary intellectual responsibility**	Field exists only in 18 record with unknown and irrelevant data
	a/NR	Entry element	
	b/R	Subdivision	
	4/R	Relator code	
801/R		**Originating source**	Fixed data ($aPT $bBN) not relevant to the aim of field. Data of field in automatically added to field 000.
	a/NR	Country	
	b/NR	Agency	
922/NR		**Type of material**	Field exists in 50 records and contains code for bibliographic level m – monographies. Code is by default added to all record in subfield 001c.
	a/NR		

APPENDIX F: Definition for using NSB/NSE

First case
If the string begins with *A, An, Das, De, Der, Des, Die, I, Il, La, Le, Les, The*, *Un* put NSB at the beginning of the string and NSE just before the second word in the string.

Example:

200 0_ aThe reality effect in the writing of history

200 0_ a≠The ≠reality effect in the writing of history

Second case
If the string starts with word which begins with L' (for example. L'evolution), put NSB at the beginning of the string and NSE just after '. There is no space before or after NSE.

Example:

2251 aL'evolution de l'humanite

2251 a#L'#evolution de l'humanite