3   FORMAT STRUCTURE

3.1   General Structure

UNIMARC is a specific implementation of ISO 2709, an international standard that specifies the structure of records containing bibliographic data. It specifies that every bibliographic record prepared for exchange conforming to the standard must consist of:

- a RECORD LABEL consisting of 24 characters,

- a DIRECTORY consisting of a 3-digit tag of each data field, along with its length and its starting character position relative to the first data field,

- DATA FIELDS of variable length, each separated by a field separator,

with the following layout:

RECORD LABEL	DIRECTORY	DATA FIELDS	R/T

R/T = Record Terminator

ISO 2709 further specifies that the data in fields may optionally be preceded by indicators and subdivided into subfields. UNIMARC, as an implementation, uses the following specific options allowed under ISO 2709.

3.2   Record Label

ISO 2709 prescribes that each record start with a 24-character Record Label. This contains data relating to the structure of the record, which are defined within the standard ISO 2709, and several data elements that are defined for this particular implementation of ISO 2709. These implementation-defined data elements relate to the type of record, its bibliographic level and position in a hierarchy of levels, the degree of completeness of the record and the use or otherwise of ISBD or ISBD-based rules in the preparation of the record. The data elements in the Record Label are required primarily to process the record and are intended only indirectly for use in identifying the bibliographic item itself.

3.3   Directory

Following the Record Label is the Directory. Each entry in the Directory consists of three parts: a 3-digit numeric tag, a 4-digit number indicating the length of the data field and a 5-digit number indicating the starting character position. No further characters are permitted in a Directory entry. The Directory layout is as follows:

Directory entry 1   Directory entry 2   Other directory entries

Tag	Length of Field	Starting Position		....	F/T

F/T = Field Terminator

The second segment of the Directory entry gives the number of characters in that field. This includes all characters: indicators, subfield identifiers, textual or coded data and the end of field marker. The length of field is followed by the starting character position of the field relative to

the first character position of the variable field portion of the record. The first character of the first variable field is character position 0. The position of character position 0 within the whole record is given in character positions 12-16 of the Record Label.

The tag is 3 characters long, the 'length of the data' fills 4 characters and the 'starting character position' fills 5 characters. After all of the 12-character directory entries corresponding to each data field in the record, the directory is terminated by the end of field marker IS2 of ISO 646 (1/14 on the 7-bit code table). For an example of a directory illustrating its position in relation to data fields see the complete examples in Appendix L. The directory entries should be ordered by the first digit of the tag, and it is recommended that order by complete tag be used where possible. The data fields themselves do not have a required order as their positions are completely specified through the directory.

3.4   Variable Fields

The variable length data fields follow the directory and generally contain bibliographic as opposed to processing data.

Data (Control) Field (00-) layout:

Data	F/T

Data Field (01- to 999) layout:

Indicators   Subfield Identifier   Other Subfields

Ind 1	Ind 2	$a (etc.)	Data	Data	....	F/T

Tags are not carried in the data fields but appear only in the directory, except for tags which are in embedded fields (see 4 - - block). Fields with the tag value 00- (e.g. 001) consist only of the data and an end of field character. Other data fields consist of two indicators followed by any number of subfields. Each subfield begins with a subfield identifier that is composed of a subfield delimiter, ISl (1/15 of ISO 646), and a subfield code (one alphabetic or numeric character) to identify the subfield. The subfield identifiers are followed by coded or textual data of any length unless stated otherwise in the description of the field. The final subfield in the field is terminated by the end of field character IS2 (1/14 of ISO 646). The last character of data in the record is followed as usual by the end of field character IS2 which in this instance is followed by the end of record character IS3 (1/13 of ISO 646).

3.5   Mandatory Fields

3.5.1   Fields Which are Mandatory in All Records

001   RECORD IDENTIFIER

100   GENERAL PROCESSING DATA

200   TITLE AND STATEMENT OF RESPONSIBILITY ($a title proper is the only mandatory subfield)

801   ORIGINATING SOURCE FIELD

3.5.2 Fields Which are Mandatory for Certain Types of Record

Records with Text

101 LANGUAGE OF THE WORK

Cartographic Records

120 CODED DATA FIELD: CARTOGRAPHIC MATERIALS – GENERAL

123 CODED DATA FIELD: CARTOGRAPHIC MATERIALS – SCALE AND CO-ORDINATES

206 MATERIAL SPECIFIC AREA: CARTOGRAPHIC MATERIALS MATHEMATICAL DATA

Electronic Resources Records

230 MATERIAL SPECIFIC AREA: ELECTRONIC RESOURCE CHARACTERISTICS

304 NOTES PERTAINING TO TITLE AND STATEMENT OF RESPONSIBILITY

Music etc. Records

125 CODED DATA FIELD: SOUND RECORDINGS AND PRINTED MUSIC

3.5.3 Mandatory Fields and Retrospective Conversion

Where earlier records do not have sufficient information for mandatory fields a default value should be used where possible. Where this is not possible the fill character should be used when allowed. As a last resort, the field should be omitted and the fact noted in the Documentation to Accompany Exchange Records (see Appendix K).

Example:

101 |#$aund

The conversion program cannot create the correct Language of the Work details.

3.6 Length of Records

The length of records, which is limited by the format to 99,999 characters, is a matter of agreement between parties to an exchange.

3.7 Record Linking

In practice there are situations when it may be desirable to make a link from one bibliographic entity to another. To give two examples: when a record describes a translation, a link may be made to the record that describes the original; or a link may be made between records relating to different titles of a continuing resource when a change of name occurs. A technique is provided in UNIMARC for making these links. A block of fields (the 4-- block) is reserved for this purpose and more information can be found at the description of those fields and in the introduction to the 4-block.

A linking field will include descriptive information concerning the other item with or without information pointing to a separate record that describes the item. A linking field is composed of subfields, each of which contains a UNIMARC field made up of tag, indicators, and field content including subfield markers. Note that these embedded fields are not accessible through

the Directory, since only the entire linking field has a directory entry. The tag of the linking field denotes the relationship of the item identified within it to the item for which the record is being made.

3.8   Character Sets

For data interchange in UNIMARC, ISO character set standards should be used. The record label, directory, indicators, subfield identifiers, and code values specified in this document should be encoded using the control functions and graphic characters of ISO 646 (IRV), which is considered the default set for the record. The code extension techniques specified in ISO 2022 are used when multiple sets are required in a record. Character positions 26-29 and 30-33 of subfield $a in field 100 are used to designate the default and additional graphic character sets used in the record. Character sets should be those established or registered by ISO but may also be the subject of agreement by parties to an exchange.

The control functions of ISO 646 are permitted in the UNIMARC record and the following are always used:

> ISl of ISO 646 (position 1/15 in the 7-bit code table): the first character of the two-character subfield identifier.

> IS2 of ISO 646 (position 1/14 in the 7-bit code table): field separator, found at the end of the directory and each data field.

> IS3 of ISO 646 (position 1/13 in the 7-bit code table): record separator, found at the end of each record.

When additional character sets are needed, the control function ESC of ISO 646 is frequently used. Two control functions from ISO 6630 used for sorting are also allowed in UNIMARC data. Appendix J gives more information on character sets used with UNIMARC.

3.9   Repetition of Data

There are four possible situations where data could be repeated in different forms:

Data appear in both coded and textual, display and non-display forms. Where possible both forms of data should appear in the record even if the information is held only once in the source format.

The document contains the same information in different languages. The International Standard Bibliographic Descriptions specify when and how parallel data should be transcribed from the item. This is catered for in UNIMARC by the use of different or repeated subfields. For examples, see field 200.

There is more than one language of cataloguing for a multilingual audience. The use of more than one language of cataloguing in, say, notes fields, is useful and in some cases mandatory within a domestic format. For international exchange purposes this facility is less acceptable: unless a receiving agency caters for the same languages as those of the source format it will need to strip out all languages except one. For that reason each record on a UNIMARC exchange tape should have only one language of cataloguing, other languages being catered for by separate records or even separate exchange tapes.

The same information is repeated in different scripts to cater for variations of sophistication of output. Ideally a catalogue entry should record a document using the script of the document. This is not always possible. For that reason, agencies with the facilities should be able to record

both original and transliterated versions in the same catalogue entry to allow the selection of the best possible option by receiving agencies. The mechanism is described in section 3.11 below.

3.10 Numerical Subfields

The following numerical subfields have been defined for use in the bibliographic format:

Subfield	Definition	Note
$0	Bibliographic Record Identifier	4-- block
$1	Linking data	4--block (Embedded fields technique); 604
$2	Fingerprint System Code	012
$2	Source	1-- block
$2	System Code	0-- block; 5-- block, 6-- block; 801; 886
$2	Link text	856
$3	Authority Record Number	6-- block, 7-- block, 500
$4	Relator Code	7-- block
$5	Institution or copy identification to which field applies: Used only for copy specific information	2-- block, 3-- block, 5—block
$6	Interfield Linking Data	
$7	Alphabet/Script of field	
$8	Reserved for Materials Specified	3-- block; 852
$9	Local use	

3.11 Treatment of Different Scripts

Record alternative graphic representations/scripts in fields 001-099 and 200-899 using content designators appropriate to the data being recorded. All UNIMARC fields will be considered repeatable for recording alternative graphic representations or scripts whether or not so listed in the body of the text. Those fields listed as not repeatable should be used no more than once per alternative graphic representation/script included in the record.

This technique is intended to provide a mechanism for recording romanizations, transliterations and alternative scripts or orthographies prepared by the cataloguing agency according to standard tables, rules, guidelines etc.

In each field repeated for the purpose of recording an alternative graphic representation/script, include both subfield $6 (Interfield Linking Data) and, if appropriate, subfield $7 (Alphabet/Script of Field). Specific instruction for the use of $6 and $7 are given below in 3.13.

3.12 Copy Specific Data

A bibliographic record may contain copy specific data in fields 012, 141, 316, 317, 318, 481, 482, 560, 600, 601, 602, 606, 610, 621, 702, 712 and 722. Fields may be linked to indicate to which specific copy a specific instance of a field relates.

3.13 Interfield Linking Data

$6 Interfield Linking Data

This subfield contains information allowing the field to be linked for processing purposes to other fields in the record. The subfield also contains a code indicating the reason for the link. The first two elements in the subfield (character positions 0-2) must always be present when the subfield is used; the third element (character positions 3-5)

is optional. Thus the length of this subfield may be either 3 or 6 characters. Subfield $6 should be the first subfield in the field (unless it is preceded by $3 Authority Record Number). It should precede any $7. Note, however, that if the alternative script representations differ also in language from their corresponding headings, then this parallel data should reside in an authority file; alternatively, mutually agreed local fields should be used by participating agencies (EX 5). Not repeatable.

Data entered in subfield $6 is recorded as follows:

Name of Data Element	Number of Characters	Character Positions
Linking explanation code	1	0
Linking number	2	l-2
Tag of linked field	3	3-5

$6/0   Linking explanation code

This code specifies the reason for the interfield linkage. The following values are defined:

a = link to alternative graphic representation/script
b = link associated with a copy
z = other reason for linking

$6/1-2 Linking number

This two-digit number is carried in subfield $6 of each of the fields to be linked together. Its function is to permit matching of linking fields and is not intended in any way to act as a sequence or site number. The linking number may be assigned at random as long as the numbers assigned to each of the fields in the pair or group to be linked together are identical and differ from the number assigned to any other pair (EX 1,2,4) or group (EX 3) within the record.

$6/3-5 Tag of linked field

This element consists of the three-character UNIMARC tag of the field being linked to. The element is optional: if the tags of both linked fields are identical, it would usually be omitted.

$7   Alphabet/Script of Field

This subfield contains the code for the alphabet and/or script for the chief contents of the field. Code values are those defined for field 100 character positions 34-35 Script of title. This subfield would usually be omitted in those fields with the same alphabet/script as that coded in 100 character positions 34-35. This subfield should be placed directly before the first data subfield (e.g. $a) of the field in which it is carried. It will usually follow a subfield $6 unless no parallel field exists, in which case there will be no $6. Following the provisions of ISO 2022 Section 1, which states that "The [character set] codes ... are designed to be used for data that is processed sequentially in a forward direction", it is assumed that characters are input in logical order. Where data, such as Arabic or Hebrew, is input in an order that supposes that it will be read right-to-left, this is indicated by '/r' after the code (EX 4). ISO 2022 Section 1 also states that "Use of these codes in strings of data which are processed in some other

way, or which are included in data formatted for fixed-length record processing, may have undesired results or may require additional special treatment to ensure correct interpretation".

It is to be understood that indication of script in $7 cannot be used as an instruction for the system. To provide correct and unambiguous interpretation of script by the system the mechanism described in Appendix J should be used. Optional. Not repeatable.

Examples

EX 1: 100 ##$a character positions 34-35 = ba [Latin]
   600 #0$6a01$a [Person as subject in romanized form]
   600 #0$6a01$7ea$a[Person as subject in Chinese script]
   700 #0$6a02$a [Person with primary responsibility in romanized form]
   700 #0$6a02$7ea$a[Person with primary responsibility in Chinese script]
   702 #0$6a03$a [Person with secondary responsibility in romanized form]
   702 #0$6a03$7ea$a[Person with secondary responsibility in Chinese script]

   Three sets of two parallel fields containing the romanized and Chinese forms of the names of the persons. The first field in each case lacks a $7 because it is in the same alphabet as that coded in 100. The linking numbers follow in sequence, although they could be in random order.

EX 2: 100 ##$a character positions 34-35 = ka [Korean]
   200 1#$6a01$a[Title in Korean characters]
   200 1#$6a01$7ba$a[Title romanized]

   Two parallel title fields containing Korean and romanized versions of the title. The first field lacks a $7 because it is in the same alphabet as that coded in 100.

EX 3: 100 ##$a character positions 34-35 = db [Japanese kanji]
   701 #0$6a04$a[First joint author in kanji]
   701 #0$6a04$7dc$a[First joint author in kana]
   701 #0$6a04$7ba$a[First joint author romanized]
   701 #0$6a08$a[Second joint author in kanji]
   701 #0$6a08$7dc$a[Second joint author in kana]
   701 #0$6a08$7ba$a[Second joint author romanized]

   Added entry fields for two joint authors, each recorded in Japanese kanji, Japanese kana and in romanized form. The fields recorded in kanji contain no subfield $7 because field 100 shows that kanji is the script of title. The linking numbers have been assigned at random.

EX 4: 100 ##$a character positions 34-35 = ba [Latin]
   700 #0$6a03$a[Romanized author]
   700 #0$6a03$7ha/r$a[Author in Hebrew. Name reads right-to-left]

EX 5: 100 ##$a character positions 26-29 = 0102, character positions 34-35 = ba [Latin]
   710 02$6a05791$aNational Library of Russia
   791 02$6a05710$7ca$aРоссийская национальная библиотека

   The Russian form of name (in Cyrillic) has been added using a local field.

Since character positions 26-29 of 100$a contain code 02 (basic Cyrillic set), there is no need to use the shift-mechanism in 791$a.

EX 6:   100 ##$a character positions 26-29 = 0103, character positions 30-31 = 05, character positions 34-35 = ga [Greek]
200 1#$a01/11 07/13Ελληνικά01/11 07/14 Windows XP01/11 07/13 - ο εύκολος τρόπος01/11 07/14

Title in Greek includes embedded phrase in English. Indicating script in $7 would not provide appropriate instructions for the system to handle combination of scripts in the field. Mechanism described in ISO 2022 should be used.

ISO 5426 (extended Latin set) has been designated the G1 set, ISO 5428 (Greek set) has been designated the G2 set. The G2 set is invoked into columns 10-15 when needed using LS2R (bit combination 01/11 07/13), displacing the G1 set. At the end of the use of the Greek set in the field the default G1 set is restored to columns 10-15 using LS1R (bit combination 01/11 07/14).