ARCA

Libraries Programme Project LIB-ARCA/2-3039
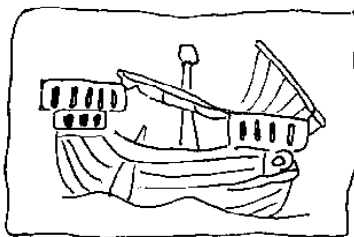
Title	:	Analysis of Bib1 Mapping
Document Reference	:	ARCA/T11/ABD
Version	:	Version 2.1
Date	:	29 September 1995
Author(s)	:	Maria Bruna Baldacci, Stefania Biagioni (CNR-IEI),
		Silvana Agueci, Fabiola Fazi, Barbara Pistoia, Renato Tamburrini (Universita' di Pisa),
		Gian Luigi Betti (Regione Toscana),
		Carmen Gomez, Jesus Zubiaga, Sara Sanchez (Sancho el Sabio),
		Ernesto Garcia Camarero, Pilar Fernandez (SABINI)
Distribution	:	CNR-CNUCE/IEI, SABINI, Fundacion Sancho El Sabio Universita' di Pisa, Regione Toscana

Abstract	:	This document deals with ARCA Target requirements concerning checking RPN queries and translating them into the OPAC language, record syntax, composition of the records and character set

Keywords	:	Bib-1 Set, Attributes mapping, RPN query processing
Reviewed by	:	ARCA Consortium members
Approved by	:	ARCA Partners

Document Status Sheet

Issue	Changes	Date	Reason
1.0 Draft A	New document	April 1995	First draft for general review
2	Major revisions	September 1995	Including requirements taking account of the Z39.50 Protocol
2.1	Editing	September 1995	
2.2	Editing	September 1996	Revision Bib1 mapping

Table of Contents

1.   Introduction

This document deals with the requirements of the ARCA Target as for bibliographic processing, and it is a complement of the ARCA Target System Requirements Document [ARCA/T11/SRD]. The present document, however, has been completed a lot of time later, after ARCA Consortium took the decision to implement some features of the Z39.50 Version 3 Protocol (which is a fully compatible superset of SR [ANSI/NISO 1995]), therefore it makes reference also to such a Protocol, whereas the ARCA Target System Requirements Document does not.

The decision to make reference to the Z39.50 V3 and to extend the requirements stated in [ARCA/T11/SRD] has been taken by ARCA Partners because users and librarians have been aware of the limitations in the user-system communication offered by the SR protocol and are mostly interested in services such as EXPLAIN and SCAN which have already been standardised by the ANSI Protocol. A similar reasoning can also be found in [PARAGON 95]. As for the ARCA Project, it has been noted that one of the components of the ARCA Target System, i.e., the Dictionary, is well suited to support the EXPLAIN facility, therefore the implementation of such a facility has been decided.

The most part of the present document is devoted to the issue of translating search queries coming from origins into the language of the OPAC to which queries are addressed. This issue will be referred to as "bib-1 mapping" because search queries are composed by a search term and by some attributes which describe the characteristics of such a term, and these attributes are specified in an official "Object" named "Attribute Set bib-1".

2.   Requirements of bibliographic processing in the ARCA Target

Requirements of bibliographic processing in ARCA regard the definition of the access points in terms of bib-1 attributes, the preferred record syntax, the composition of the records, and the character set.

2.1 Mapping onto bib-1 set

According to the ARCA Technical Annex, the most meaningful aspect of bibliographic processing in the ARCA Target is the check of the queries coming from the origins (i.e., whether queries are acceptable or not by the OPAC) and their translation into the language of the interfaced OPAC.

The basic structure of a query encoded in the SR search APDU is composed by a list of attributes, or *attribute combination*, followed by a *term*. There are six types of attributes: Use, Relation, Position, Structure, Truncation and Completeness, which can also be referred to as type-1 to type-6 attributes. The Use attribute (or type-1 attribute), if provided, identifies the access point against which the *term* is to be matched. The Relation, Completeness, Truncation and Position attributes, if provided, specify additional match criteria. The Structure attribute, if provided, identifies the form in which the term has been supplied. No other information is given about the term, for example, about its language or whether it is case sensitive or not.

To check whether the query is acceptable by the OPAC, the ARCA Target must know which attributes combinations can be accepted by the OPAC.   The operation of determining which attributes combinations are acceptable can be seen as a mapping of OPAC access points and access modes onto bib-1 attributes combinations. The information resulting from bib-1 mapping, useful for query control, is to be stored in the ARCA Dictionary [ARCA/T11/SRD]. Because of being ARCA portable on different OPACs, information necessary to the above operations is different for each different OPAC, and is to be loaded into the ARCA Dictionary at the Target system starting up, after librarians of that OPAC have defined it.

The attribute set bib-1 is presented in the ISO SR Protocol without any guideline [see Annex 1], therefore the work by ARCA librarians was initially affected by much uncertainty, especially about the meaning of attribute of type 3 (Position), 4 (Structure) and 6 (Completeness); much help has been given by the document prepared by the Z39.50 Implementor Group [Attribute, 1995]. In the following a brief report and comments on the work done are given.

2.1.1.Extending the SR attribute set bib-1.

The bib-1 document prepared by the Z39.50 Implementor Group (ZIG) [see Annex 2] adds many new attributes to those listed in the SR bib-1 Set. Of these attributes, some of the Use type were used by ARCA librarians because they capture the different meaning of "author" versus "name" (Use attributes 1004, 1005, 1006, 1009) ; others because they map OPAC access points for which no Use attribute was found in the SR bib-1; for example, the Use attributes 1018--publisher and 1021--bibliographic level.

 Among the attributes added to type-2 to type-5 ones, only the Structure attribute 101 (name, normalized) has been used: this attribute is well   suited to the searching on OPACs authority files, when present.

As mentioned above, there is no attribute to specify the language of the *term* in the RPN query, whereas such an attribute would be necessary   in searching by subject on a multilingual OPAC as SABINI is. For this reason, a special attribute will be used in the ARCA origin to target communication and the Z39.50 Implementors Group will be solicited to consider such a question.

No use attribute for "printer" nor for "place of publication" has been found: these access points are very useful in searching ancient materials.

2.1.1.2.   Selectors

An argument not mentioned in either protocol is how to manage *selectors*, i.e. those data elements that can only be used for reducing result set obtained during user-OPAC interaction.

As selectors are not access points, the OPAC should not accept Search APDUs queries containing one or more Type-1 attributes mappable on OPAC selectors. However, the ARCA Target can be made to check the query to verify whether Type-1 attributes mappable on OPAC selectors are ANDed with other Type-1 attributes mappable onto OPAC access points: if they are, the ARCA Target can process the query in order to translate it in a form acceptable by the OPAC. The solution given by the ARCA Target is described in [ARCA/T12/ADD].

2.1.2.Attribute combinations

To verify whether a Search APDU query can be accepted by the OPAC, the ARCA Target has to check each combination as a whole, and not its single attributes; this depends on that a given attribute combination may not be accepted by an OPAC even though the OPAC is able to process its single attributes. For example, an OPAC might be able to process the Structure attribute "word list" but only when combined with the Use attribute "title" or "corporate author", and not, for example, when combined with "name, personal" or with the Position attribute "first in field"; moreover not all non-Use attributes combinations can be accepted, even though the single attributes are; for example, the combination of attribute "word list" with "first in field" seems to have no meaning, therefore should not to accepted even if the single attributes are applicable.

Within an attribute combination, each attribute type is optional. Traditional OPACs are searchable only through their access points, thus the presence of a Use attribute is mandatory in queries directed to them; however most recent OPACs are capable to match search queries against the contents of the whole records, thus no access point need to be specified.

The protocol does not address Target behaviour when one or more attribute types are not supplied in a query. A given Target might supply a default attribute, dynamically select an appropriate attribute based on the other attributes supplied (for example, the Position attribute can be given the value "Any position in the field" if a Title WORD search is received, or the value "First in field" if a Title PHRASE search is received) or fail the search because it requires that the attribute type be supplied.

This opens the door to different political choices regarding discouraging or encouraging Targets to provide all attribute types and Origins to use all attributes to precisely specify user needs. The option of encouraging seems to be preferable in the context of   very large OPACs in order to avoid noise. But this contrasts, for example, with the choice

made by the Library of Congress WWW/Z39.50 gateway which only provides the two attributes Use and Structure and default values for the other four.

In the ARCA librarians' opinion, however, this issue needs further discussion, after experience has been made; moreover it presents different aspects depending on whether the SCAN Service , or the EXPLAIN facility are available,   because this can strongly affect the precision of the query and, consequently, the precision of the response.

As a first attempt to map the query languages onto bib-1, ARCA librarians have used all the attributes of types 1 to 5 when applicable. No Type-6 attribute (Completeness) was used in the mapping, because this Type has been considered to be too restrictive. Targets could provide the default "value -1--incomplete subfield" or ignore it at all.

The work done by ARCA librarians concerned the OPAC of the Region of Tuscany (CUT-ISIS, for short) and the OPAC of the Sancho El Sabio Foundation (SES-SABINI). Description of the two OPACs is in Sect.3. The results of the work made for mapping attributes are shown in the Table I, II, III.

Tables I and II sketch, for CUT-ISIS and SES-SABINI, respectively, which are the bib-1 attributes each OPAC can support. The tables are organised in six columns; the first one lists the access point names of the OPAC; each access point is followed, in the second column of the same row, by the corresponding type-1 value(s), and, in the third to sixth columns, by the Type-2 to Type-5 attributes each access point supports. Each value in any column may combine with all the other values in the same row. Each row concisely represents, therefore, all the attributes combinations possible for a given access point (in some cases, however, combinations for a given access point span over more than one row, for avoiding forbidden combinations).

CUT-ISIS is also capable to support attribute combinations where no Use value is present, or where the Use attribute 1016---Any (or 1035---Anywhere) is present. The desirability of this is however to be verified and therefore such attributes have not been presented in Table I.

As an example of how supported attribute combinations are translated into the OPAC's query language, Table III shows specimen "AttributesPlusTerm"'s for searching by title and names (left side of the Table), and, correspondingly, their translation into the query language of CUT-ISIS (right side of the Table). It can be noted that some use attributes of the bib-1 set cover more than one access point of CUT-ISIS; see, for example, that the terms qualified as "1--Name-personal", are matched against two access points of CUT-ISIS, because in such an OPAC personal names are accessible through different points according to whether they refer to a "Main author" or to an "Added author".

Table I: CUT-ISIS

Denomination	type 1: use values	type 2: relation values	type 3: positional values	type 4: term structure values	type 5: truncation values
Title	4,34	3	3	2,6	1
Publisher	1018	3	3	2	1
Title-series	5	3	3	2, 6	1
Note	63	3	3	1, 2	1
ISBN	7	3	3	2	1
ISSN	8	3	3	2	1
Author-personal	1003,1004	3	1,2	1,2	1
		3	-	102	-
Author-corporate	1003, 1005, 1006	3	1,2	1,2	1
		3	-	102	-
Title-uniform	6	3	1,2	1,2	1
Author-pers. added	1003,1004	3	1,2	1,2	1
		3	-	102	-
Author-corp. added	1003, 1005, 1006	3	1,2	1,2	1
		3	-	102	-
Subject	21-29	3	1,2	1,2	1
		3	3	2,6	1
CDD	13	3	1	2	1
Number-local-call	53	3	1	2	1
Abstract	62	3	3	1, 2	1
Code language	54	3	1	2	-
ATS	1003	3	-	102	-

Table II: SES-SABINI

Denomination	type 1: use values	type 2: relation values	type 3: positional values	type 4: term structure values	type 5: truncation values
Personal author	1, 1004	3	1	1,2,101	1, 2, 3
		3	3	2,6	1, 2, 3
Personal name-subject	1, 1009	3	1	1,2,101	1, 2, 3
		3	3	2,6	1, 2, 3
Personal name-series	1, 1004	3	1	1,2,101	1, 2, 3
		3	3	2,6	1, 2, 3
Corporate author	2, 1005	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Corp.name-subject	2	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Corp.name-series	2, 1005	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Conf. name-author	3, 1006	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Conf. name-subject	3	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Conf. name-series	3, 1006	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Title	4,35-44	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Title series	5	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Uniform title	6	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Uniform heading	6	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Uniform title subject	6	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Uniform title series	5, 6	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3

table follows in the next page

Subject	21, 29	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Name-geogr.-subject	58	3	1	1,2	1, 2, 3
		3	3	2,6	1, 2, 3
Place-publication	59	3	1	1, 2	1, 2, 3
		3	3	2,6	1,2,3
Publisher	1018	3	1	1, 2	1,2,3
		3	3	2,6	1,2,3
Date-printing	?	3	1,3	1,2	1
Classification-CDU	14	3	1	1	1
ISBN	7	3	1	2	1,3
ISSN	8	3	1	2	1
Number-legal-dep	49	3	1	1	1,2
Number-local-call	53	3	1	1	1
Accession-number	12	3	1	2	
*Date-publication	31	2, 3, 4	1,2,3	4	
*Code language	54	3	1	2	
*Code-geogr-area	55	3	1	2	

*   selector

Table III: Mapping into CUT-ISIS query language

AttributesPlusTerm	CUT-ISIS Queries
1-4 2-3 3-1 4-1 <LIBRARY AUTOMATION>	LIBRARY AUTOMATION/(1)
1-4 2-3 3-1 4-1 5-1 <LIBRARY AUTOM>	LIBRARY AUTOM$/(1)
1-4 2-3 3-1 4-2 <LIBRARY>	LIBRARY/(1)
1-4 2-3 3-1 4-2 5-1<LIBRAR>	LIBRAR$/(1)
1-4 2-3 3-3 4-6 <LIBRARY AUTOMATION>	LIBRARY/(1) * AUTOMATION/(1)
1-4 2-3 3-3 4-6 5-1 <LIBRAR AUTOM>	LIBRAR$/(1) * AUTOM$/(1)
1-1 2-3 3-1 4-101 <ROSSI, MARIO>	(ROSSI/(9) * MARIO/(9)) + (ROSSI/(12) * MARIO/(12))
1-1004 2-3 3-1 4-101 <ROSSI, MARIO>	ROSSI/(9) * MARIO/(9)
1-1004 2-3 3-1 4-2 <ROSSI>	ROSSI/(9)
1-2 2-3 3-1 4-1 <REGIONE TOSCANA>	REGIONE TOSCANA/(10) + REGIONE TOSCANA/(13)
1-2 2-3 3-3 4-2 <TOSCANA>	TOSCANA/(10) + TOSCANA/(13)
1-1005 2-3 3-1 4-1 <REGIONE TOSCANA>	REGIONE TOSCANA/(10)
1-1005 2-3 3-3 4-6 <REGIONE TOSCANA>	REGIONE/(10) * TOSCANA/(10)

2.1.2.1. The ATS-1 Profile

ATS-1 [see Annex 3] is a Profile approved by the Open Systems Environment Workshop-Special Interest Group on Library Automation (OIW SIG/LA). This Profile specifies the use of attribute set bib-1 within a Z39.50 RPN query for searching by author, title, or subject. Its purpose is "to ensure that complying origins and targets can provide basic search access to bibliographic data bases, similar to the common online catalog systems used in many libraries"; because of this, the profile is a very restrictive one.

If considered from the point of view of the ARCA OPACs, the main faults of the ATS-1 Profile are that it does not allow for the well established practice of distinguishing between "Personal author" and "Corporate author", nor for searching titles, subjects and corporate authors by word or word list, nor for limiting search results by *selectors* (date, language, etc.). However ARCA librarians consider a standard Profile to be necessary, because the reliability of Z39.50 search results is questioned by the large number of attributes listed in bib-1.

For the above reasons, all OPACs should implement the ATS-1 profile and should be able to activate it when an origin declares to conform to it. As for the OPACs experimenting the ARCA Target, the ATS-1 profile will be enriched at least with entry points "Personal author", "Corporate Author" and "ISBN/ISSN" in the experimentation phase.It remains to decide how much to extend this basic search profile, or, conversely, which subset of the Tables I and II is to be loaded in the ARCA Target Dictionary for the operating phase. This will be decided after evaluating the results of experimentation.

2.1.3. Comments

When reading the document on the semantics of bib-1 prepared by the Z30.50 implementors group, librarians acquainted with US MARC tags can better understand the meaning of Use attributes because each attribute is   made to correspond to one or more US MARC tags. Other librarians might like some attributes be specified in more details or referred to UNIMARC tags.

The interpretation of the following attributes has been a bit laborious:

1-57---name and title: it is a derived key used in indexes;

4-6---word list: the interpretation of its meaning is problematic because it is not inferable whether words in the list: are to be considered as ANDed or ORed . In our context, words have been considered as ANDed.

"phrase" versus "string": from reading comments in the document "bib-1 semantics " of september 1995, the difference between them are not clear. Generally OPACs search by matching strings rather than by checking that all the words in the "term" appear, for example, in the title with the same order; however in the ATS-1 profile the Structure attribute "phrase" is mandatory for title searching, and this is to be clarified.

2.2.   Preferred record syntax

Preferred transfer record syntax is UNIMARC for both SES-SABINI and CUT-ISIS. SES-SABINI already exports UNIMARC records. CUT-ISIS presently exports records formatted according to ISO 2709 and will provide their conversion to UNIMARC.

Annex C of ISO-10163 also includes the SBN-record syntax among the preferred ones. However, the SBN record syntax is used by the Italian National Bibliographic Service only for very specialised data bases. Therefore it has not been considered as a possible transfer syntax for the OPAC of the Tuscany Region.

2.3 Record composition

SES and CUT export full bibliographic records and short ones as well. The minimum composition of short records, expressed in Unimarc Tags, should be as follows:

Tag 200: Title and responsibility area

Tag 205: Edition area

Tag 210: Publication area

Tag 215: Physical description

Tag 010/011: ISBN/ISSN

Tag.990: call number / serial holding

2.4 Character sets

SABINI represents bibliographic information with the following character sets:

ISO 646 - Unaccented Latin base set

ISO 5426 - Latin alphabet expanded

ISO 6937 - Coded graphic character set for text communication (contains diacritical marks together with characters and ligatures used in western European languages) [1]

ISIS is a multilingual system, and its character sets include also non-latin alphabets.

UNIMARC characters are coded according to ASCII expanded codes, which has been internationally standardised as ISO 646.

Admitted character sets are the following:

ISO 5426 - Latin alphabet expanded

ISO/DIS 6937/2 - Added set.

In the UNIMARC record directory, only ISO 646 is used. Information about which type of sets are used is registered in the UNIMARC field 100, subfield $a.

---

[1] According to the information maintained in the ECHO WWW, ISO 6937 is being superseded by ISO 8859 (accented characters used in European Languages) and by ISO 10646 (Universal Multiple-octet Coded Character set.

Theoretically, the systems SABINI and ISIS should be able to communicate with each other and   with all the OPACs capable of processing UNIMARC. ARCA librarians' intention, however, is to discuss issues related to communication between OPAC's using different character sets with their Greek colleagues involved in the   HELEN CEC Project.

It is worth noting that the Z39.50/SR protocol operates in a 8-bit environment and that BER encoding requires that.

3 Description of the OPACs of Region of Tuscany and of Sancho el Sabio Foundation

3.1 CUT-ISIS OPAC

CUT/ISIS is the union catalog of the libraries in the Tuscany Region. The catalog contains ca. 400.000 items and is distributed on CD-ROM so that all regional libraries can search it locally; moreover an experimental version, named TECA.CD, is accessible through INTERNET at the address http://vik.regione.toscana.it The construction of a networked CUT accessible through the SR Protocol, which will act as the Tuscan node of the Italian National Library Service (S.B.N.), is one of the objectives of the ARCA Project.

The CUT OPAC is constructed on the generalised Information Storage and Retrieval System CDS/ISIS [CDS/ISIS, 1992]. Such a software, freely distributed by UNESCO, has been designed specifically for the computerised management of non-numerical data and runs on MS/DOS and on UNIX platform as well.

CDS/ISIS is able to manipulate an unlimited number of data bases, each of which may consist of completely different data elements. Moreover, it allows selective creation of inverted files for each database, extracting fields, subfields or elements thereof. Postings in the inverted file contain additional information precisely identifying the field from which the data was extracted, as well as the relative word position within the field.

CDS/ISIS supports boolean query language (operators AND, OR, NOT, PROXIMITY) and right truncated search terms. It is capable of importing as well as exporting records structured according to the Standard ISO 2709. Linguistic versions exist for all European and for most non-European languages, including Arabic, Chinese and Korean.

The CUT catalog, which has been realised with the CDS/ISIS software, allows searching through the access points listed in Table I together with the bib-1 attributes supported by each of them.

3.2 SABINI-OPAC at Sancho El Sabio Foundation

Sancho El Sabio Foundation is a private centre, entirely subsidised by the Caja de Ahorros de Vitoria y Alava, aimed at the conservation and diffusion of Basque bibliographical heritage. Its OPAC, which allows access to the bibliographic descriptions of books and serials, is constructed on the library automation system SABINI [SABINI, 1994].

SABINI manages one or more catalogs, each consisting of the master file of bibliographic records and a lot of auxiliary files where access point values are maintained. The set of possible access points is shown in Table II; however the access point set can be tailored according to the needs of different installations. For example, an access point "Author", not comparing in Table I, could be defined by grouping access points "Personal Author", "Corporate Author, and "Conference Name Author".

SABINI allows search, retrieval and browsing facilities through its user interface and through its command language as well. The basic search command has the following structure:

Search, access point, term.

Successful searches result in a result set identifier followed by the number of items retrieved. The result set identifier can be referenced in subsequent commands for i) displaying bibliographic records, ii) limiting the result set through the values of selectors (Language, Year, Nation), iii) combining the result set in boolean sentences using operators AND, OR, NAND.

SABINI is capable of importing as well as exporting records with the following record syntaxes: UNIMARC, USMARC or IBERMARC.

Table II lists the access points allowed by SABINI together with the corresponding bib-1 attributes.

References

[ANSI/NISO, 1995]   *Information Retrieval: Application Service Definition and Protocol Specification*. ANSI/NISO Z39.50-1995. Available: ftp.loc.gov.

[ARCA/T11/SRD]   *ARCA Target System Requirements*, ARCA/T11/SRD, 15 July, 1995.

[ARCA/T12/ADD]   *ARCA Architectural Design Document*. In progress.

[ARCA/T21/SRD]   *User Interface Application Requirements Document*, ARCA/T21/SRD, 12 June 1995.

[ATTRIBUTE, 1995]   *Attribute Set Bib-1 (Z39.50--1995): Semantics*. September 1995. http://lcweb.loc.gov/z3950/agency

[CDS/ISIS, 1992]   *CDS/ISIS : versione 3.0 per mini e microcomputer : manuale d'uso* . Traduzione italiana coordinata e rivista da Giampaolo del Bigio. - Firenze : Titivillus, 1992. - ISBN 88-7218-005-8.

[ISO, 1992b]   International Organisation for Standardization (ISO), *Information and Documentation - Open Systems Interconnection —Search and Retrieve Protocol Specification*. ISO 10163-1:1993(E)

[PARAGON, 1995]   PARAGON, Project LIB-SR-TARGET/2-3034. *WP200-Task 230: Target Requirements Analysis*. 20 May 1995.

[SABINI, 1994]   *SABINI : Automatizacion de Bibliotecas : Manual de Uso*. Madrid : Sabini, 1994.