

Standards for Serial Holdings and for Serials Data in the Serials Analytic Record

by Alan Hopkinson
Middlesex University

1. Introduction

This study is undertaken within the context of MODELS, Moving to Distributed Environments for Library Services. The standards for serials holdings and their implementation in existing automated systems are considered not in the abstract but within the context of their usefulness in the distributed library environment.

Most materials required by library users fall into two categories, the monograph and the journal article. As far as requiring assistance in finding it, the journal article takes far more time to locate but for the researcher it is generally the more crucial.

Journals are now being made available in electronic form, but only a limited number of titles, and for the future there will be the need for retrieval of older material which may never be converted, so it is essential that a method be established for identifying uniquely serial holdings in printed form. Indeed, many of the characteristics of a serial article are not present in the electronic serial, and the article in the electronic journal is no different from an electronic monograph in the way one locates it.

Many users retrieve journal articles through services such as First Search or BIDS. Also they find them through references in monographic or periodical literature that they are reading. Their next step is to acquire the appropriate article.

To take an example of a 'normal' print journal: the ideal would be for the reader to be taken directly from a search in an abstracting and indexing service to their institution's catalogue. The system would search the catalogue for the title, and then the particular issue, and would also determine the availability of that issue. If the issue were in the collection but not available at that time, the reader could be informed and a reservation placed. If it were not available, the system would search outside the library, perhaps in libraries of the reader's choice within a subset determined by his library. An option could be provided then for the journal to be provided on inter-library loan or for a photocopy of the article to be made. Any messaging used for the retrieval of the issue should therefore retain the reference to the article originally required in case it can be satisfied.

There has therefore to be a matching exercise between the journal article and the library holdings. The ideal should be to provide an exact match which will identify the record of the required item in a collection and examine it to see the current availability of the item.

Many managers and developers of library systems are interested in implementing this kind of matching but are deterred from setting up the required mechanisms because the standards necessary to implement them are not yet determined.

However, the metadata we are looking at to a greater and lesser extent are already present. Therein may lie one of the problems for the adaptation of existing systems, since the metadata may yet be present but not in a format satisfactory for the types of operation alluded to above to be performed satisfactorily.

This study deals with the object of a search being a serial part (volume or issue); other studies are covering the result of a search when it is the record of an article.

1.1 Definitions

In the remainder of this paper we will use the term serial for journal as this is the form used in standards and is more inclusive, including as it does journals, newspapers, periodicals and even, in some cases, monographic series, though the definition which follows excludes monographic series.

The British Standard definition (see for example BS 5999¹) is:

"A publication, in printed form or not, issued in successive parts usually having numerical or chronological designations and intended to be continued indefinitely." The definition continues: "Serials include periodicals, newspapers, annuals (reports, yearbooks, directories, etc.), the journals, memoirs proceedings, transactions, etc. of societies and monographic series. This definition does not include works produced in parts for a period pre-determined as finite."

Another major characteristic of a serial is that it is not an intellectual entity: each serial and in most cases each part contains multiple articles which are intellectually speaking monographs in their own right and sometimes are even bound as monographs, particularly in the case of those serials which have parts published on a particular topic and by one author.

However, the serial as a whole, though not of intellectual importance has physical predominance. Libraries obviously arrange serials as serials not by their contents and indeed catalogue the serials rather than catalogue the articles.

There is even a tendency for library catalogues to record the serials and make little or no mention of the individual parts despite their importance for the user of the library. This survey is concerned therefore at the level at which serial holdings are recorded in catalogue and the standards used for recording them.

The term 'holdings' in this survey refers to the metadata representing those parts of the serial which have been acquired by the library or libraries covered by the metadata.

This survey will use the term parts to mean parts of a serial published individually which may or may not be bound up into volumes which are also parts.

This survey will ignore the problems of contributions in monographs which, it cannot be denied, have many of the characteristics of serial articles, but which overall are a vastly smaller proportion of items required by library users. It is a reasonable assumption that pagination would usually be all that is required for articles in monographs. Pagination is a requirement of retrieval of journal articles so if solutions were developed to enable the location of journal articles there will be no residual problems for monograph articles.

2. Existing metadata

2.1 Source data for the article

The source data from the secondary service may not necessarily be complete but if it has come from a contemporary service it is likely to be, otherwise the service would go out of business in the competitive market place. Readers may nevertheless come up with references gleaned from books, articles or reading lists which are lacking in specificity and may want to search on that data. All things being considered, data for an article should include, as well as the author and title of the journal article, the pagination of the article in the serial, the part and volume enumeration, the title of the serial which may or may not include a 'series', and must be complete when it includes in its title bibliographic levels, such as Journal of Physics Part B. The title may be qualified by a unique identifier of which the current leader in the field is the ISSN², an alternative or additional being the CODEN³. A convenient name for the data relating to the location of the article in the source serial is 'host statement'. This term is used in ISBD⁴ and is less open to ambiguity than 'source'.

2.1.1 Source data from secondary services

There is quite a variety of formats for data in secondary services whether they are accessed by CD-ROM or on-line. Since the data are serving a very specific need, it is surprising there are so many variations on a theme given that there must be some conformity in order that the resulting records be comprehensible and able to serve the functions for which they are intended, in particular the need to match a reference with a serial part. Nevertheless, the relationship between volume and part, number or whatever one wishes to call it is a topic that in the 1970s exercised the committee that maintained the UNISIST Reference Manual⁵ and the group that developed the UNESCO Common Communication Format (CCF)⁶. Both these exchange formats took into account the needs of secondary services, attempting to give serial articles equal weight with monographic material. They were being developed at around the time the secondary services were getting off the ground and the need for standardisation for the purpose at least of producing a model to prevent re-inventing the wheel was strongly felt. The Reference Manual received input from a group of secondary service data providers called the Four Ways Group which consisted of representatives of ASIDIC (Association of Information Distributing Centers), EUSIDIC (European Association of Scientific Information Dissemination Centres), NFAIS (National Federation of Abstracting and Indexing Services) and ICSU/AB

(International Council of Scientific Unions/Abstracting Board)⁷. It is therefore useful to see what resulted. There was certainly in the minds of the developers of both formats a feeling that there was a need to cut through the terminology popularly applied (i.e. the captions) and define the elements in a neutral and unambiguous way. The Reference Manual has two separate fields labelled 'Volume Identification Data (First Order designation)' and 'Issue Identification data (Second Order Designation)'. Volume has four subfields, 'Caption', 'Number', 'Year' and 'Other identification of part of volume'; issue identification has the same but without a year. Date, which in the deliberations again consumed a great deal of time, was felt to be the date of publication and only belonged with volume when the serial needed a volume number but had nothing but year that could be supplied. At any rate the presence of subfields indicates the need to retrieve the separate elements so treated. As with date, pagination caused certain problems; did pagination belong to the volume or to the article? Could the same field used for pagination of a monograph be used for pagination of a serial article? The Reference Manual had a separate field for each.

The CCF group was unable to be so specific in this data element. It was devised as a format which aimed to cater for the lowest common denominator and had to provide a carrier which could contain both data from the library world in the UNIMARC format⁸ and data from secondary services in the Reference Manual format; thus, it placed data relating to first and second levels and to pagination in separate subfields of one field.

The subfields are:

- A Volume/part numeration and designation
- B Pagination defining a part
- C Other data defining a part

The main reason for the loss of the specificity found in the Reference Manual was the lack of specificity in the UNIMARC and other MARC formats which the CCF had to take into account. Certainly, many users of MARC formats informally used the serial holdings field for the volume and issue of a serial article. The author of this study did this when designing the database for IDIS at the Institute of Development Studies because in UNIMARC, the format used, there was nowhere else for it to go. When later they had to reformat their data for a secondary service, the unformatted field with volume and part data was the most difficult to deal with. Standards have been needed here for a long time but has still not been agreed on.

The root of the problem is that publications are not all identical in their structure and that publishers do not always number their serials, even those with identical structures, in the same way.

Parts may be called parts, numbers and issues. This leads to a desire to accommodate a flexibility which is unnecessary for the end user of a system.

Newspapers are often seen by their publishers and also by the creators of secondary services metadata as having numbers and no volumes even if volumes exist. All data

providers seem to standardise their captions. Most include captions, some use punctuation.

Of the databases investigated, many of the structures of the data were identical because they used the same software as each other for their search engine and database. Also, database publishers probably impose a certain level of standardisation on their database producers.

Wilson Indexes

Wilson indexes have a host statement which begins SOURCE and is followed by the journal title and ISSN in parentheses. This is then followed by volume which may be preceded by v {space} or ns (without following space) for new series. If there is a volume, number is not usually included. For dailies and weeklies, volume is not included. Number is included preceded by no (abbreviation for number) without a space. Finally the date appears, in English, with the month in full but the year abbreviated as '95 and the day of the month where applicable following the month (e.g. February 2 '95).

If the date on the part is not sufficient to define the part uniquely, the part numbering is included. Practice seems to be to avoid the part and prefer the precise date when available.

Sight & Sound (ISSN:0037-6702) v 1 p 44 January '92
Art monthly (ISSN:0142-6702) no199 p21-2 September '96
Cahiers du Cinema (ISSN:0008-011X) no458 p 80-1 July/August '92
Mnemosyne (ISSN:0026-7074) v 43 fasc1/2 p132-49 '90
Architects' Journal (ISSN:0003-8466) v 201 p16-17 February 2 '95

Bowker Saur indexes

There are slight differences in practice between indexes. In Applied Social Sciences Index and Abstracts, the host statement begins with the journal title and is followed by the volume number without any caption where there is one, otherwise by the date in parentheses. Dates are abbreviated to last two digits and months are abbreviated in English. Where there is a volume the date is included follows by the date in parentheses or the part number in parentheses. If the date is included as well when it is not absolutely necessary to include it for unique identification, it is not then included with parentheses. ISSN and CODEN come after the pagination. Note that practice may change slightly over time.

New Law Journal 146 8 Mar 96 p.348-9

Presumably continuous pagination mean that the date is not 'needed'.

Spectator 266 (29 Jun 91), p.9-11
Nursing Times-Nursing Mirror (14-20 Jan 87), p.46-7

British Humanities Index

Bookseller (4680) 1 Sep 95, p.34-65
Greece and Rome 38 (Apr 91), p.82-186
Journal of Garden History 16 (1) Jan-Mar 96, p.1-66
Book Collector 43 (4) Winter 94, p.529-48

It will be noted that the date is always included but where it is thought to be not necessary for the unique identification of the part it is not in parentheses.

ABI/INFORM

ABI/INFORM is clear because each element is identified by a caption.

The volume, part and date are always included where they exist and are preceded by Vol: Iss: Date: The date is in American order with the year in full Feb 24, 1997 with a comma only where the day is present and would otherwise run into the month with only a space between.

Journal: Economist [ECT] ISSN:0013-0613 Vol: 342 Iss: 8001 Date Jan 25, 1997
p:70

Journal: Business Week [BWE] ISSN: 0739-8395 Iss: 3405 Date: Dec 9, 1996

IBSS

IBSS is produced by the British Library of Political and Economic Science (the Library of the London School of Economics), is available through BIDS and it has some characteristics of a library catalogue (see also 2.1.2 below).

The host statement is preceded by JN and consists of title, date with month abbreviated, volume and part preceded by captions) and pagination. Volume numbers are sometimes in Roman numerals presumably based on the source serial. In the last example below, only the date is provided which serves alone to identify the part.

Volkerrechte, 1995, Vol.48, pp.19-77
Middle East Studies Association bulletin, Dec 1995, Vol.29, No.2, p.240
Orita, Jun-Dec 1995, Vol.XXVII, No.1-2, p.127
Administration, Summer 1996, Vol.44, No.2, pp.159-176
Quarterly bulletin. Central Bank of Ireland, Winter 1996, pp.39-48
Social Sciences Citation Index

Social Sciences Citation Index when accessed through BIDS has a similar host statement.

Modern Asian Studies, 1994, Vol.28, No.Pt3, pp.533-556
Modern Asian Studies, 1987, Vol.21, No.JUL, pp.511-519

In this example can be seen a difference which could amount to a change in practice over the years or merely be an inconsistency.

2.1.2 Source data from library catalogues

Some library catalogues contain records of serial articles which have been thought by the library staff to be useful. In some cases the article may have been photocopied in which case it may well have attained to all intents and purposes within the framework of the catalogue and the library holdings the status of a monograph. These records usually stand as individual entities. Often when records are converted (for example when a new system is procured) they become anomalies, requiring special treatment. The main reason for this is that neither UK MARC⁹ nor USMARC¹⁰ provides a standard way for treating them. The data identifying the containing volume and part is sometimes placed in the MARC field for serial holdings because this has the same pattern (though the field is usually treated as free-form). Thus, in some databases, field 255 (Numeric and/or alphabetic, chronological or other designation area (series)) may contain data relating to one part e.g. Vol.95,no.4(April 1990) rather than what the UK MARC manual states, data relating to the journal run which would be

Vol. 1, no. 1 (Jan.-Feb. 1977)-

or in the case of a completed journal

Vol. 1 no. 1 (Nov. 1943)-v. 10, no. 12 (June 1953)

Since UK MARC is not intended for the holding of journal issue data this is non-standard.

In any case, generally journal article data will be in a variety of formats.

Many citations do not include the ISSN. They begin with the title of the serial, sometimes abbreviated, sometimes in full, though with 'and' and its equivalents replaced by '&'.

They are usually followed by volume and part/issue number and spanning pagination. Volume and part may be preceded by Vol. and no. and variations on these, but it is probably fair to say that the captions should be taken from the document. The date may come before the volume number or after the part number. There is a feeling that there should be standardisation of the terms equivalent to volume and part, and this is important in the context of databases with material from serials originating in different languages where captions are different.

2.2 'Holdings' data for serials in automated catalogues

The holdings data against which the serial articles (or the requirement of a user which may be more inclusive than a single article) are to be matched exist but in a less consistent form between systems.

There are a number of problems inherent in recording serials holdings, which result from the very nature of the material. The main problem is the levels into which a serial is divided. Usually there are two which are generically called volume and issue in the English language. If there is only one level as in the case of newspapers they are often called number rather than issue. In other languages there are different terms. Given a system with suitable identifiers such as punctuation used in a standard way, it would be possible to dispense with language-based captions. Nevertheless some automated systems encourage the entering of the data without captions but add standard captions (e.g. vol., v. no.) for the convenience of the user ('no.' is utilized here as a generic term). Some journals use the year as a volume number in which case Vol 1997 is not appropriate. Nevertheless the year is clearly used as such when each year's parts start at one but there is no volume designation. This system begins to fall down with more complex journals. There are few journals with more than two levels. But some journals start renumbering from Vol. 1, sometimes after a gap such as the kind of gap that took place in the publication of some learned journals in the Second World War. Then there will be vols 1 to (say) 40, followed by New Series, vols 1 - 90. Additional complexities are language editions or editions in different media.

Historically serials have been catalogued separately from monographs. Recently there have been moves to incorporate serials in the opac. Arguably this may have been fostered by the introduction of automated serials checkin systems into integrated library automation systems. At the start of the library automation era, it was common to utilise separate serials control systems. Some, such as Blackwells ISIS were provided by serials vendors and were intended to facilitate the ordering and acquisition of serials from the usual suppliers. Desiring to increase their market and following the trend particularly noticeable in the UK towards totally integrated library automation systems, the general system suppliers incorporated serial 'checkin' systems (the name by which they are commonly known) into their integrated packages. Sometimes these systems were developed in conjunction with the supplier of serials; no doubt they always reflected the functionality of pre-existing serial systems. These systems are primarily concerned with the initial acquisition of serial parts. Some systems regularly conflate them into a holdings statement. Some systems allow the last so many parts received or due to be displayed in the opac. The concept of 'date due' is very important as the systems mostly concentrate on prediction of the date of receipt of forthcoming parts so that claims can be generated to the publisher. For this to work, the systems need to take account of all possible frequencies and they usually include codes to represent these. This area of speciality does not really interest us although the serials checkin system may dictate the display of serial holdings in the catalogue and the format of the data immediately behind the display format.

These automated systems can of course cope with multiple copies some of which copies will of course find their way to different branch libraries. Some systems have incorporated binding data but few systems include circulation data, though this is not

difficult in a relational database management system as it requires only extra fields for loan status, borrower identifier, date of issue and date of return.

There are also instances of checkin systems which do not interact with the rest of the system, and to which the library reader does not have any direct access.

There has always been the capability for serials to be included in the standard library catalogue, but there has never been the mechanism for ensuring the kind of interchange metadata required for electronic data exchange. Catalogues which include serials will obviously include serial titles. If the data are derived from external sources, they will most probably have the frequency and a starting date originating from a BNB record. They will also have notes fields and BLCMP records which have included additional fields from USMARC include related titles in the form required for an access point in their records. Such records will also include ISSN. Holdings must be added locally. They may not have been added with any consistency except in those organisations which are part of union catalogues. Moreover any holdings statements may not have been added with any real consistency. The Anglo-American Cataloguing Rules (AACR)¹¹ have little to say about serials holdings: the code contains the following examples in 12.7B20 Copy being described, library's holdings and restrictions on use.

Library lacks: Vol 12, v.16

Library has v.1, 3-5 and 7

These are very sketchy; the code does not even state when the cataloguer should list the lacks or the holdings though common sense would tell the pre-automation cataloguer, that the holdings that are almost complete should list the lacunae and the holdings which are far from complete should list the parts that are actually held.

Since the majority of libraries have used AACR (or rules based on AACR) even for serials, there has been little externally-imposed consistency on the format of holdings other than this. Moreover, the above rule was only introduced into the second edition (1988 revision)¹² of AACR2, they are not there in the second edition. Given that some existing union serials records in catalogues are based on records which were entered as MARC records before the revision of AACR2, one could expect an even lower level of consistency (no examples were given in AACR 2 [1978]).

Interestingly, the London Union List of Periodicals¹³ which is based on a Libertas system and which imports on a monthly basis data from member libraries has a display very similar to AACR 2 prescriptions. The consistency achieved there is generated because the system makes it easy for staff to enter volume and part numbers in a consistent form and the system adds consistently the captions which, if AACR was followed to the letter, should be taken from the document but clearly are not in this example.

See below for a discussion of ISO 10324¹⁴ which is an extension of the rules found in AACR but is probably not yet implemented anywhere in the UK.

2.3 Holdings data for the serial in CHECKIN systems

Most serials checkin systems have metadata for an issue of a serial held on the database up to but not including its loan status. This data may not interface with the cataloguing system though in SIRSI's Unicorn for example there is a facility to display in the OPAC the information of the last serial parts however much is needed. Checkin systems are primarily concerned with the receipt and prediction for claim of predicted but non-receipted parts. Some systems now include a program to collapse the data relating to the parts into a holdings statement which may include more than one range.

Horizon for example includes the following data elements to cope with this in the holding_summary table: applicability date, gap, start enumeration, start date, start chronological pattern, end enumeration, end date, end chronological pattern, free enumeration, free chronology, note. These data are generated from data which treat each part on its own and indicate its availability in terms of whether the library has yet acquired the part. These data are usually called serials checkin data. On the basis that the system does not expect serial parts to be loaned, the capability for circulation of serial parts has not been built into Horizon. However, systems could extend this data with availability data; the status code could be included to indicate on loan as well as available, and new fields could be added for date of loan, date due back etc.

Horizon, Fretwell-Downing and UNICORN have a mechanism for transferring serials checkin data to the MARC record and no doubt other systems can do this. Records contain a field which can be updated by running a program to collate the data from the serials checkin data into a holdings statement and also provides details on the last few holdings.

Most checkin systems appear to have a mechanism for recording data from different levels of inclusiveness. Compatibility with the USMARC Holdings Format requires this. However, it is very difficult in the context of a study of this kind to discover about the way data are stored inside systems and the possibilities for output, other than by what can be seen from viewing the systems. Customers of automated systems and potential customers have a hold on system suppliers in the way that participants in research activities do not; sales or public relations staff usually handle this kind of request for information and the information they give may be no more accurate than a visual inspection of the system displays.

2.4 Holdings data: a summary

Clearly, the best source of holdings data in integrated library automation systems is the checkin systems which have a record for every part in the serial. In practice, many systems do not have this data. Those that have serials holdings statements automatically compacted from checkin data supplemented by statements produced

manually may have indications of the runs available with or without mention of omissions. It is not usually possible to tell from sight of these statements what the level of accuracy is. An interesting question is whether, to make holdings statements explicit, and to make them matchable against requests for parts, it is necessary for the publication pattern to be available. If a journal is published monthly except August and if July is called issue 7/8, or if the numeration goes up to 11 rather than 12 does this matter for retrieval. One can assume that the onus is on the searcher to ensure that requests are made only for parts which exist, and if a request is made for a part which does not exist it is satisfactory for a response to be given that that issue is not available. Publication patterns are notoriously complex and also they change. It is probably best to set up systems that avoid the need for knowledge of publication patterns at the time of the search for a particular issue.

Clearly this area of library cataloguing has suffered because of the lack of an accepted standard. Nothing is going to advance until standards are accepted.

A practical problem of some significance in opening up many existing sources of serials data in academic library catalogues is the rough and ready way in which serials have been treated in the past. The production of a catalogue which would reflect precisely the situation on the shelves was never felt necessary because the library user or member of library staff on behalf of an external requestor would go to the shelves to see what was actually there. It is only when we consider a MODELS type of environment that this method becomes intolerable.

Source data from the past are seldom complete. Checkin systems exist primarily to control the acquisition of journal activities and not their subsequent activity. The data therefore have been created for this purpose, and may lack data complete enough for their use for circulation. The only way to rectify that is manual.

Thus it is evident that even if the standards proposed as essential to enable matching between requests and the data in library catalogues become available and accepted from tomorrow, there will be databases which have been developed before these standards which will need to be converted.

Each data source will need to investigate whether its collection is valuable enough to make it worthwhile to upgrade its data, and if so, how best it can achieve these standards, by writing conversion programs, by manual exercise or no doubt a combination of both. Obviously there are financial implications here.

The next section deals with those standards that are available though not necessarily used.

3. Available standards

Section 2 has covered the results of data relating to serials metadata being incorporated into systems without the backing of reliable standards covering all the areas of interest.

This section covers the standards that are available at present.

Librarians are quite strong on standardisation: their natural inclination is to look for standards and follow them to a greater or lesser degree (where they exist), though perhaps they are as guilty as any other profession at not adopting the latest version. This is due in part to the difficulty of upgrading large quantities of data which requires usually manual re-cataloguing.

3.1 Standardisation in recording serial parts data

Standardisation in the recording of serial parts requires standards for three components: the name of the title, any further identification for the title and an agreed representation of volume and part numbering. The first two are being dealt with elsewhere.

3.1.1 SICI

SICI is the Serial Item and Contribution identifier and the standard document is available on the Internet (<http://sunsite.Berkeley.EDU/SICI>) as a ballot draft¹⁵.

It is to be used with serial publications in all formats. It is intended to be used with EDI, SISAC bar codes, Z39.50 queries, URNs, electronic mail and 'human transcription in print'.

The SICI is divided into three sections, Item segment ('item' has the same meaning of 'part' in this document), Contribution Segment and Control Segment

The item is the physical part which contains the journal and it includes ISSN, chronology and enumeration

The contribution segment includes pagination and a title code (if the SICI is identifying a part - i.e. a volume or an issue - and not an article this will be null; this is Code Structure 1). Contribution segments may be defined 'locally' (this is Code Structure 2).

This does not concern us in this study as the aim of the study is to investigate the matching between the serial holdings and the SICI, for which the article identifier is not relevant. The third segment, the control segment, includes an identifier of the code structure, a derivative part identifier, a medium/format identifier a standard version number (indicating which version of SICI is used) and a check character.

Alphabetical characters are always in upper case. In addition to digits 0 to 9 there is a limited character set allowed.

The first 9 characters are the ISSN of the journal followed by a chronological designation followed by the enumeration. The date in standard ISO date format¹⁶ (YYYYMMDD) is as specific as is required going down to the day of the month if necessary. Although it is dealing with a single part it may be necessary to use a spanning date for parts which have a start and end date and this is denoted by a slash. The repeated number always starts from the year. Thus we can have (199502/199503); this is February / March 1995. Parentheses must be present even if no date is available (seldom the case) to separate the ISSN from the enumeration. Codes beginning with 2 replace the month for the seasons and 3 for the quarters.

A plus + after the closing parentheses indicates supplement and asterisk * an index. These are used when they are published separately, not when they included as a physical 'contribution'.

The hierarchies of the enumeration are separated by colons.

The contribution segment is enclosed in < >. <> must be present if null. The control segment follows. Neither contribution nor control segment need concern us; they are described fully in the standard.

The SICI is highly dependent on an ISSN and on such having been assigned to the lowest level necessary and then used consistently.

The Library (Sixth Series) XVI:4 (Dec. 1994)
is represented by SICI:
0024-2160(199412)6:16:4

IFIP Transactions A (Computer Science and Technology) A:7 (1992) becomes SICI:
0926-5473(1992)A:7

Institute of Electronics, Information and Communication Engineers Transactions A,
March 1994, vol. J77-A, (no. 3) becomes 0913-5707(199403)J77A:3

There ought to be no room for ambiguity but there could be confusion in the minds of the users where a journal is split into 'parts' and each has its own ISSN. The 'parts' should then logically be at the serial entity level and the volumes of the 'parts' at level 1, a different situation from that of IFIP Transactions A above.

In general the SICI seems a suitable identifier as it is both meaningful to the human eye and capable of being parsed by computer. It should pose no problems in the description of the majority of serials.

As it stands, it can be used only for a single item and not for a run (except for such examples as a volume of individual parts).

The SICI has no place for publication pattern.

The SICI has been added to the list of Z39.50¹⁷ use attributes as item 1037 SICI. It appears in the approved extensions to bib-1 attribute set published by the Z39.50 agency in April 1997¹⁸.

3.2 Standardisation in recording serial holdings data

3.2.1 MARC

MARC is a good example of a standard which is not a standard. Most systems today have chosen to adopt the MARC 'standard' for the exchange of metadata between systems. On the one hand, MARC has never been made a formal national (or international) standard; only the record structure is a standard ISO 2709¹⁹. It must not be forgotten that there may in any particular application still be slight divergences from any given MARC standard justified on various grounds which may have a greater or lesser air of credence. Also there are a number of different MARC standards in use in the British Isles such as UK MARC, BLCMP MARC²⁰, USMARC and OCLC MARC²¹. UK MARC is regarded by many as the national standard for the exchange of bibliographic data but in truth it is the British Library standard for the production of the British National Bibliography and a few other things beside and as such does not cover many of the data elements required for serials. Perhaps following from the librarian's reliance on standardization the logical corollary can be seen taking effect: if there is not a standard for something, then little effort seems to be made to do a good job in that area. The area is dealt with in an ad hoc way, and of course different libraries do their own thing.

There are two areas in MARC which are relevant. The first is the data in the MARC serials record. In UK MARC there is one format which covers both books and serials, in USMARC there are separate formats for bibliographic, authority and holdings records. To avoid confusion between different practices that may be prevalent in the UK, this study will confine itself to UK MARC where there is anything applicable in UK MARC; in the absence of anything in UK MARC, it will look at what USMARC has to offer.

3.2.1.1 MARC: Holdings field 255

There are no instructions (other than what could be deduced from the examples) for using field 255 Numeric and/or alphabetical, chronological or other designation area (series). This field applies to serials as well as series. Multivolume monographs which do not concern us in this study are dealt with in field 248. There is one subfield only into which all data goes. This subfield is repeatable though there are no rules about when it is to be repeated. On the grounds that any rule is better than none, it could be decided that the subfield be repeated for distinct 'runs'. In example 6 in the UK MARC manual, there is a serial which must have restarted under the same title but with new numbering (my spacing); the new numbering has a single level hierarchy.

255 00 \$aVol. 1, no.1 (Nov. 1943)-v. 10,no.12 (June 1953); No. 1 (July 1974)-

This seems an unlikely example for the run of a serial with a large gap in publication and for the serial to be regarded as the same serial after such a long break!

The enumeration again follows fairly closely AACR 2 (1988)

3.2.1.2 MARC: Numbering and chronological designation note: 515

The examples given for this field are concerned with the enumeration and not the holdings themselves.

3.2.1.3 USMARC Holdings fields

The USMARC Holdings format²² covers a multitude of different circumstances. It is primarily a means of exchanging data on holdings in general, which can be holdings represented by a union catalogue, a single systems' catalogue with multiple sites or a single site library contributing to a union catalogue.

In this study only the features applicable to serials concern us.

When the requirements for serials holdings are studied it is clear that many different conditions have to be catered for. For example, if parts of a serial have been bound into volumes, it is necessary to find some way of stating that fact for the purposes of loan or inter-library loan.

The format, in line with other USMARC formats, includes a coded field (008) which is different from the other formats' 008 fields. Because this field is different from other 008 fields, records in this format could not easily be processed by computer programs which have been developed for processing records of books or serials as complete entities. However, that does not matter because the format allows all the fields enumerated below to be included in a normal MARC record (so that they could be added to UK MARC if required). Given that the practice in the UK has always been to have one integrated format (even to the extent of including authority records in the MARC bibliographic record for the convenience of exchange) it is necessary only to include the fields that would be added to the MARC record and not be concerned with those fields such as the fields for linking to the related bibliographic record which need to be only used if the standard is to have a separate holdings record. (These linking fields are intended to link to the record number in various numbering systems: field 004; to standard number (LC Control Number); field 010, Linkage Number (related internal record control number from another system): field 014; ISBN: field 020; ISSN: field 022. Additionally, field 024 covers International Standard Recording Code, Universal Product Code, International Standard Music Number, International Article Number.

The fields specifically for Holdings Data are as follows:

- 853 Captions and Pattern: Basic Bibliographic Unit
- 854 Captions and Pattern: Supplementary Material
- 855 Captions and Pattern: Indexes
- 863 Enumeration and Chronology: Basic Bibliographic Unit
- 864 Enumeration and Chronology: Supplementary Material
- 865 Enumeration and Chronology: Indexes
- 866 Textual Holdings: Basic Bibliographic Unit
- 867 Textual Holdings: Supplementary Material
- 868 Textual Holdings: Indexes
- 876 Item information: Basic Bibliographic Unit
- 877 Item information: Supplementary Material
- 878 Item information: Indexes

As can be seen, the format is all-inclusive: it seems to have covered all possibilities. The record can deal separately with the basic subject of the holdings, the bibliographic unit as well as its supplements and indexes.

The first in each set of these fields need only concern us by way of example, though the second and third are equally applicable to serials and would have to be included if this were adopted as part of a MARC standard.

Fields 876-878 contain data relating to individual copies and their circulation status. These fields would of course be necessary to make the picture complete as to whether individual parts were available.

There are three groups of fields covering these bibliographic items.

Captions and pattern (853)

These fields indicate the caption and pattern of each level of enumeration of a serial.

Both MARC indicators are used in this field. The first relating to captions patterns refers to the extent to which the data have been compressed.

The second indicator identifies whether or not all levels of captions are present. There are separate subfields for levels of enumeration (down to 6!) and four levels of chronology. The second indicator shows whether these are all present if they exist.

The subfields contain codes for frequency, calendar change date and regularity pattern relating to the captions. These are outside the scope of the holdings and are required for serials checkin systems.

Enumeration and chronology (863)

These fields contain the identifiers of the parts and are the essential 'core' of the record. Each part may be retained in a separate instance of field 863; or a field 863 may contain a 'compressed' statement.

Here is an example of a complete 863 with its parallel 853 to denote captions for printing out. 853 is not required in a system where it is not desired to use captions.

```
853 20$81$av.$bno.$u4$vr$(year)$j(season)$wq$x21
863 30$81.1$a1-10$i1943-1952$zbound
```

The first indicator of field 863 relates to the levels of description in Serials holdings statements (ANSI Z39.44)²³. In this case the holdings are described to level 3 of the standard. The second indicator value of 0 indicates the data are compressed but that a readable statement may be generated from them. Value '1' would indicate no compression, each field relating to one published issue. Value '2' indicates compressed but use textual display found in field 866. Value '4' is used when uncompressed statements cannot produce a reasonable display and would point to an 86 field which would produce a holdings statement such as '1974-1981: many issues lost'.

Subfield 8, Link and Sequence Number, contains a repeated field count linking to field 853.. if the captions and enumeration change a new 853 is required denoted by its subfield 8 containing 1, 2, 3 etc. Field 863 in turn reflects this enumeration.

It can be seen from the example above that 853 controls the data as held in field 863. Subfields a and b (with the possibility of continuing to f) are successive levels of enumeration with the caption held in 853 and the enumeration in 863 in the example above compressed but with the potential for one 863 field per issue. Note that this field 863 does not include a subfield b since the holdings statement comprises only volume and not the second level. No doubt if a volume was only partially available, subfield b would be used; field 863 is set up for any journal to cover any possible combination of subfields that might be required for field 864. Subfield u in 853 indicates there are 4 of the 'numbers' to each 'volume'. It follows the enumeration for the level, indeed for each level, to which it applies and so is repeatable. It will logically never follow the top level, subfield a. \$i and \$j relate to chronology and are entered in parentheses when they are required only to interpret but are not to be displayed. \$w is a frequency code and \$w is a code to indicate when the calendar changes. The item to which this record relates is quarterly denoted by the seasons and the code 21 indicates that the volumes start with Spring.

Textual holdings (866)

Whereas the previous fields tackle the problem of holdings in a logical way to assist in machine manipulation of the data, this field deals with the text as it should be displayed in a holdings list. This field may be used only as a last resort when the data in a compressed form cannot produce a record as required by a standard dealing with visual display. The first indicator specifies the Holdings Level, the second the Type of Notation, either ANSI Z39.44 or ANSI/NISO Z39.57²⁴ or ANSI Z39.42²⁵. The subfields here are straightforward; \$a Textual holdings; \$x Non-public note; \$z Public note with \$8 Link and sequence number (as described above).

An interesting example is given of the use of this field

853 01 \$afasc.\$wx
863 42 \$81.1\$a1-29
863 42 \$81.2\$a30-40
863 42 \$81.3\$a41-124
866 41 \$81.2\$afasc. 30-40\$zsome missing

Field 853 has indicators which mean 'Cannot compress or expand' and 'Captions verified, all levels may not be present'.

This cannot be expanded because the details of fasc. 30-40 are not known and it cannot be compressed because it is already in the compressed form.

Subfield \$a indicates the first level of enumeration fasc.: there are no more in this example. \$w indicates the frequency.

In the next three fields, the holdings have been split up into three groups. The 863 fields are linked by their code in \$8 to the 866 fields and the 853 field. Since 853 applies to them all it does not need a \$8. 866 only applies to the middle group of 863 fields. The idea is that 866 would replace the middle 863 in an eye-readable catalogue; presumably there is no code available in the original source to indicate 'some missing'.

Here MARC is behaving as an exchange format which has to be able to carry data from other sources with less (or sometimes more) specific data.

This always has to be a consideration in systems. Let us imagine a scenario where it was decided to adopt something like USMARC for Holdings in UK Academic libraries. Some databases would not be able to provide anything like the specificity outlined in the example above, offering nothing more than the information that the serial was present in some form or another in the library. Certain library databases may have checkin systems which could provide this kind of specificity, but the chances are that no libraries could have such specificity for their older material. Checkin systems do not usually have the data relating to serial issues loaded retrospectively and as most checkin systems have been around for only about 10 years it is unlikely that older material would be included. Usually the best that happens when a checkin system is installed is that earlier material is held in a general statement in the system for the sake of completeness, but this is likely to take the form: Vol 32, no. 1 - (some issues missing).

Item information (876)

Item information relates to an individual item and includes subfields for Internal item number (e.g. automated system related number since an accession number goes in 'piece designation'), Invalid or cancelled internal item number, cost, date acquired, source of acquisition, use restrictions, item status, e.g. lost or withdrawn, temporary location, piece designation (i.e. a barcode), invalid or cancelled piece designation, copy number, non-public note, public note, materials specified and link and sequence number.

The format document describes how a holdings record may be embedded in the record of a serial using the record embedding technique of USMARC.

3.2.1.4 Relationship to eye-readable standards

It is possible to produce a holdings statement following the appropriate standard valid in the US, Serials Holdings Statements (ANSI Z39.44). ANSI Z39.44 does not recognise the concept of subfields but has data elements separated by punctuation.

The predecessor of ANSI Z39.44 was ANSI Z39.42, Serial holdings statements at the summary level. This has been adapted by ISO as an International Standard which is at present in its final stages before publication, ISO 10324 Holdings statements : summary level. This is discussed elsewhere in this report.

3.2.1.5. Overview of USMARC

The format is quite detailed in its specifications but stops short of indicating the status of an item, whether it is available or not. We will deal here with the holdings fields embedded into MARC bibliographic record (i.e. a MARC serial record).

Subfield 876 \$j is for Item status but is described as containing information about the permanent status of the item. \$l is for temporary location and states 'This subfield is not intended to provide circulation information.' This would have to be added as a private subfield or field. Much of the format is intended to provide the standard holdings statements for display. It therefore concerns itself with compression. It has to be borne in mind that the format is intended for the exchange of data between systems in the way that MARC systems conceived it, batch transmissions of data by tape. The document states that the standard is for 'those involved in the communication and processing of MARC holdings information'. However, as with all MARC formats the tape standards are referred to and in common with other MARC formats it can be used to transmit data on-line and no doubt is, though the standard mentions only transmission by tape. That is one reason why it does not concern itself with temporary status, i.e. on loan, but only permanent status such as lost or missing. Detailed though the format is, it is a batch updating tool. It can provide a mechanism for the updating of a union catalogue. It can provide a model for standardisation of the data elements for an internal system whether it be an individual catalogue or a union catalogue. But it is not a suitable mechanism for enquiring of another database if a particular item is present or not. A different standard is required for the content of a message relating to the presence or otherwise of an individual part. That standard, of course, has to be compatible since a database that can provide data on serials parts should be compatible with MARC standards and with a standard for enquiring about a particular part.

What is needed is a mechanism of reporting the situation as it stands at any given moment in time. This requires a method of searching holdings data dynamically when the request for the record is made and producing the data in the MARC record that is

returned in response to the Z39.50 query. For this we need the holdings field for the holdings of the required part (which could, indeed, be based on the format of 863 used for uncompressed data) These data are not usually accessible from the cataloguing module in current library automation systems but the presence or otherwise of individual parts is usually available in the checkin module. Of course the whereabouts of individual journal parts as far as loan is concerned is not recorded in a systematic way (for example ad hoc records may be produced at the issue desk to record their loan as they are often not regularly loaned to all categories of borrower). However, if there is to be any mileage in indicating whether these elements available to a remote user, it has got to be recorded on the local system!

So in the ideal world we need mechanism for recording all the data on a particular issue, its existence, its acquisition by the library and its loan status. This should be done by extending the uncompact 863 field. USMARC has many various ways of dealing with this field including textual notes to qualify subfields which are linked to particular fields. Many of these variations exist no doubt to enable the standard to be fully compatible with ANSI Z39.44 Serials holdings statements. If we do not have to be compatible with this eye readable format, a simplified 863 is possible.

3.2.2 UK MARC

There are no incompatibilities between USMARC and UK MARC which would get in the way of the adoption of these fields in UK MARC. There has never been any automatic adoption of USMARC tags by UK MARC. Moves were begun to harmonise UK MARC with USMARC and originally this was to be phased in by 1999. However, it has been agreed that the process of complete harmonisation will have to be phased in over a longer period²⁶.

Whenever UK tags are added, steps are taken by the British Library to ensure that there is no clash in the enumeration used, in the subfields or in the indicators, with USMARC practice. Therefore, the fields and subfields relating to holdings and location in USMARC could be added formally to the UK MARC format if required.

3.3 ISO 10324

ISO 10324: Information and documentation: Holdings statements: Summary level is an international standard at present at FDIS status, which means it is about to be published (voting finished on 19 February 1997 and a few minor amendments were considered at a meeting in London in May 1997). It has been prepared by ISO TC 46 SC9. It has been based on the ANSI (US) standard which has been developed hand in hand with the MARC format for Holdings and which is very close to BS 5999:1980 *Specification for serials holdings statements for library and documentation centres*.

The document follows the style of a cataloguing code, such as the International Standard Bibliographic Description (ISBD).

The aim of the standard is to specify 'display requirements for holdings statements at the summary level to promote consistency in the communication and exchange of holdings information.' Additionally the standard claims it is appropriate for union catalogue lists.

There are three levels of holdings statement; the first indicates only the existence of a serial title in an institution. The second adds to the first the extent of an institutions' holdings. Level three includes a summary statement.

The standard deals with 6 areas which are further subdivided.

1. Item identification is outside the scope of this study; it includes the item for the bibliographic item, title, ISSN, CODEN, etc.

2. Location data area includes the institution identifier, sublocation identifier, copy identifier (for multiple copies) and call number.

3. The third element is entitled 'Date of Report Area' and is an optional element to indicate the date of the latest modification of the record.

4. The General Holdings Area is a coded field which is essential for serial holdings. The whole is included in parentheses and the five subelements are separated by commas. If codes are not available or the data to establish the codes, a textual field may be used; then there may not be 5 data elements.

4.1 Type of unit designator indicates whether it is a basic bibliographic unit, or a supplement or an index.

4.2 Physical form means text, video, microform, etc and there are codes for these.

4.3 Thirdly there is a completeness designator which consists of 4 values appropriate for serials, 0: information not available, 1: complete (over 95% !), 2: incomplete 50%-94%; 3 very incomplete or scattered (less than 50% held).

4.4 Acquisition status designator relates to whether the item is 2: complete or ceased, 3: On order; 4: Currently received; 5 Not currently received or 0 Not available, 1 other

4.5 retention designator indicates whether the latest copies are retained, whether replaced until microform, or permanent retention.

4.6 Extent of holdings: The most important one for public identification of serial holdings is the Extent of Holdings area. The standard is very specific as to punctuation here since the punctuation allows the holdings to be expressed in a concise form.

Here are some examples of the use of the permitted items of punctuation:

| | |
|--|---|
| v.1:pt.1 units | The colon separates levels of hierarchy in the bibliographical units |
| v.1,v.3 | Comma indicates a gap in a range of holdings |
| v.1/2 1969/1970 chronological year data | Slash indicates combined numbering or combined, or non-chronological year data |
| v.2:no.5=fasc.15 | Equal sign separates alternative numbering schemes |
| v.1-3 | Hyphen indicates an unbroken range |
| <Water damaged> | Angle brackets indicate the specific extent note |
| v.1:no.3;pt.6 | Semi-colon separates two levels below the second level. |
| 1950-197? | The question mark indicates an unknown final digit of a date |
| "index" | Quotation marks enclose the name of the unit. |
| [1980/1981] | Square brackets enclose enumeration or chronology. Optionally encloses enumeration for incomplete parts.] |
| v.1(1983) | Parentheses separate enumeration and chronology when the data are recorded together |
| v.1-9 10 | The blank is used to separate data elements within an area. |

Complete examples of how these are used are given (see below).

Captions are optional and should not be used if the original document has none. If used they should be abbreviated according to the prescriptions of ISO 832 *Rules for the abbreviation of bibliographic terms*.²⁷ Most examples include captions. Are they really needed? It should be possible to extract the essential (numeric) part of the data in any computerised matching process but for many systems this may not be easy and it would probably better if the captions were not there.

Complete examples:

v.1-19+"suppl."v.1-12
v.1(1950)-10(1959)
v.1-5(1901-1905)
v.2-6,8-14,17-20, 1945-1949,1951-1957,1960-1963
v.1(1950)-2(1951),4(1953)-8(1957)
v.1(1929)-[3](1930)-8(1936)
v.5-6(1950-1951),10(1955),12(1957)

1912-1950,1954-
v1:no.1
Bd.1:T.1;Nr.3
i:2;1
ser.1:v.1-ser.3:v.25
Bd.1-
v.58(1970)
v.1-105
v.1-5
v.
v.1-3=no.1-36
ser.1;no.1-ser.1:no.4,ser.1:no.6 = no.1-16,no.21-24
1/2
1969-
Bd.1(1968)-
t.2(1940)-9(1947)
1969:Jan
v.1-6 <bound> v.7-10 <unbound>

4.7 Holdings note area. This may be used to give information on a gap or a non-gap break, or any other information, in a non-formatted form.

This standard is very much a standard for the production of eye-readable holdings statements. Its punctuation is not distinctive enough to make automatic parsing feasible. The SIC is a more suitable alternative.

3.4 Holdings of electronic serials (electronic journals)

Holdings of electronic serials may be recorded in conventional catalogues in the same way as serials in other media. The main difference would be in the location, i.e. the location would indicate they were 'electronic' rather than the shelf location. The access problems, technical and copyright are not part of this study.

Therefore any recommendations relating to holdings for hard copy journals should be applicable to electronic serials.

3.5 DTDs

3.5.1 Dublin Core

The Dublin Core Metadata element set is under development²⁸. The main aim is to describe sources of information on the Internet in such a way that search engines could be pointed to 'headers' that consisted of index data. Such index data will often be equivalent to a MARC record in their content: hence the perceived need for compatibility between the two that the Library of Congress have addressed (see 3.5.2)

Dublin Core is necessary for indexing electronic texts and texts of electronic serials. Of course, the difficulties of locating serial articles in their physical parts are not a problem with electronic serials; nor can Dublin Core help with the locating of hard copy serial issues. Thus in Dublin Core 'Source' is described as 'Objects, either print or electronic, from which this object is derived.

3.5.2. USMARC DTD

The Library of Congress is developing MARC Document Type Identifiers which are intended to be used in field testing and preliminary implementation of MARC in SGML (Standard Generalized Markup Language)²⁹ systems (<http://www.loc.gov/marc/dtd/marc/dtdback.html>)³⁰.

The file which can be downloaded includes a complete description of the MARC formats including the holdings format enabling the representation of all MARC fields in HTML.

Since the holdings fields are all present, it can be assumed that any records will be exactly equivalent to USMARC fields.

The DTD will be applicable mostly to electronic serials which will be able to have embedded within their SGML electronic indexing data which will serve to provide records for external indexes as well as the headers in the journals themselves. There has not been much experimentation with the DTD yet and the Library of Congress believe that there may not be much of a market yet but the stimulus to do something was provided by the fact that various people were doing projects in different places and could have come up with different results had there not been a lead from the MARC Office.

3.5.3 Simplified SGML for Serial Headers

Pira International developed an SGML application for serial headers on behalf of Book Industry Communications³¹. This is intended to carry data on serials in the headers of the serials or serial articles which could be used in other applications such as the production of bibliographies of serial articles. This incorporates the SICI as well as having the separate data elements of the SICI available in their own right.

4. Recommendations for further work

This section offers two alternatives, the SICI and MARC and suggestions for Z39.50.

In the context of these recommendations, a number of factors must be borne in mind.

Firstly, it is the responsibility of system providers to enable the data in their systems to be compatible with any standard for serials holdings.

Secondly, it is the responsibility of the owners of the data who wish to participate in cooperative projects to ensure that data produced in the past is sufficiently complete

and in a format which can be understood in the current system. They may do this by a number of automated methods but it is unlikely that this can be done completely without manual intervention. When most serial item data were originally entered or even when present library automation systems were planned, the kind of retrieval and data matching required now for serial issues was never imagined. Thus it would be very surprising if the data needed now were available.

Thirdly, as with any exchange format or other method for transferring data between systems, recommendations like these are not intended to dictate how a system should hold its data internally, though there clearly are implications for internal storage of data when an external requirement is very precisely stated.

4.1 SICI

Judging from its description, the SICI could, without any problems, be adopted as the format of data for the serial contribution and the serial item, as it is as an eye-readable format, its elements well-defined and uniquely identifiable by punctuation. It would include a certain element of redundancy if used as an identifier for a serial issue within a serial record because the ISSN would be present on every serial issue record as well as in the record of the journal itself. But that should not be a problem. If storage space were a consideration a SICI stripped of its ISSN with a code indicating it was known but not incorporated could be used. But matching with an incoming SICI would then become more difficult. Additionally, a SICI look-alike code for holdings should be developed for the format of data for consolidated or compact holdings. This will be more complex than the SICI itself since it will have to cope with spans. Omissions should probably not be permitted; a span would begin again after a gap.

The hyphen (-) is used only in the ISSN of a SICI, the solidus (/) being used when there is a double issue. Therefore a hyphen could be introduced to indicate spans and even open ended dates, but it should only indicate one complete span and there should be no possibility of indicating omissions. The use of this would have the benefit of consistency between single (uncompacted) records and compacted.

An example of a 'spanning SICI'. Note that this has no authority but has been devised for this study. Code 4 is used to indicate this is a spanning SICI. ? indicates code for version of SICI (this would need a new code) and X the check digit (modulus 37) which has not been calculated here.

0001-3218(197201-199212)6:1-26:11<>4.0.TX;?-X
0001-3218(199601-)-6:30:1-<>4.0.TX;?-X

Here two separate spans are included, so the SICI has to be repeated.

It would be interesting to see how successful an automatic conversion of, say, *Libertas* style holdings (where the two levels are consistently and clearly identified) to a SICI-type standard could be. Of course, when using existing data one must remember that certain data elements may not be available in the bibliographic records and there should

be a mechanism for including these as not available rather than non-existent - perhaps the bar sign could be used as in the following example where the date is not available. Date or volume/issue identifier are the only possible omissions as there is some redundancy here.

V.1; no.3- 0000-0000(||||)1:3-

The form of the SICI could also be adopted as the mandatory format for the contents of the holdings fields in MARC. This is discussed below.

4.2 MARC

MARC is important in the equation as far as a UK implementation of Z39.50 is concerned (though the situation could change) because one of the favoured options for a Z39.50 retrieval is that it should result in the return of a MARC record³². Given this scenario, the retrieval of a record with appropriate data when a serial issue is the target results in the client retrieving from the server the complete MARC record of the serial. It is not usual for serial holdings at the part level to be held in the MARC record, but there is no reason why a MARC record containing them could not be generated on the server at the request of the client system. In just the same way as a MARC record can include repeated fields for the metadata relating to the copies (by convention 949 or 966 are fields used for this: 949 is the field used by OCLC when exporting copy data to other systems); so a MARC record can include a repeated field each instance of which related to an individual part.

However, it might be more productive when looking for a serial part to be able to return a record or create on the fly a record representing the required part at the time of doing the search. There are two possibilities: the search is made for the actual part on the host system (which can result in the return of a MARC record of that issue), or the search is made in the returned MARC record after it is returned.

It can be argued that the search for the record of an individual item rather than for the entire serial record should be made as early as possible in the process. The MARC record of a serial containing the desired part should be returned with a field containing the full information on the desired part. The system would go down not only to the holdings but also to the circulation. This however is an ideal scenario. It is not realistic to suppose that there are many systems that can do this at present. There are a large number of links in the chain that need to be completed.

- a) The serial holdings metadata must be recorded somewhere in the data base!
- b) The serial holdings metadata must be accessible for the purpose for which it is required, i.e. indexed to the level of the individual part and its availability. The level of standardisation available in this area may not be sufficient to make this feasible. This must therefore be included in the profiles.

c) There must be a way of communicating the holdings to the MARC record. The alternative scenario is that the client does its search and finds the serial title which matches. The MARC record is then returned and the client does the analysis. Because of the inconsistency of recording of serial holdings (and in some cases their inherent fuzzy nature), it may at present be more feasible for the client to do the analysis and present the results to the user allowing the end user to participate (manually) in some of the necessary decision making.

Nevertheless, in order to enable systems to do this, mechanisms should be put in place to enable these data to be available where it can be provided. If the standards are not in place, no development can begin.

4.2.1 UK MARC format

For UK use, there is a great deal to be said for implementing fields from USMARC. However, USMARC holdings fields are constructed differently from the SICI. Their fields have been devised to be compatible with ANSI Z39.44 Serials holdings statement and therefore may not be completely compatible with the requirements of searching multiple library holdings. It would be difficult to insert subfield identifiers into the SICI which would be necessary to provide the data from the SICI in MARC holdings form. Then the beauty of the simplicity of the SICI would be lost. One possibility would be to have a UK MARC field for holding the SICI with two versions, one corresponding to the individual SICI which could represent either an item or an article; the other the spanning SICI proposed for spans of holdings. These different data elements could be distinguished by indicators, or by the use of different subfields or by the use of separate fields. The choice would be best made with the assistance of the British Library UK MARC Office. Repeatable fields would be used for each.

In May 1997, there had hitherto been no moves to add the SICI as a distinct field to USMARC because it can be entered as a System Control Number in field 035 but it was being considered by the USMARC Office.³³

However, it could be decided that a spanning SICI is not feasible or desirable. For example the owners of SIC could object to this use. Moreover, the SICI does not have any code for serial publication pattern whereas the MARC holdings fields do.

Therefore, if it was agreed to use the USMARC holdings format, the necessary fields and subfields should be added to UK MARC.

In that case UK MARC should adopt from USMARC a simplified 863 field which would be equivalent to the compacted USMARC 863. It might also be desirable to include field 853 for the pattern to better enable parsing of compacted holdings statements. The British Library should be approached to advise on principles involved in picking and choosing from USMARC. If it were decided that this was not possible

but the fields had to be adopted in their entirety, there is no reason why projects using this for Z39.50 should not agree that only this subset be used.

As with the SICI, this kind of uncompact statement of holdings will not be attainable for all material without recataloguing. However, there will be different levels of quality of data in different systems. In some cases only records at the serial level will exist. In other cases, records exist with holdings statements. It would be far easier for the holding library to parse these holdings and produce repeatable fields for each issue.

Some kind of collaborative effort could be set up which would involve the patterns of serials being established and exceptions being noted so that libraries with holdings statements including omissions could generate the fields that existed with their exceptions. This exercise would create large MARC records but this should not be a serious problem for present day systems and even less of a problem in the future as storage costs decrease.

Interestingly, the National Library of Australia where USMARC has been adopted have not made use of the USMARC Holdings format³⁴.

4.3 Z39.50

4.3.1 Z39.50 Supplementary Profiles

The Z39.50 standard requires supplementary profiles to enable interoperability between systems as far as the serial metadata are concerned.

These are required at the level of the elements on which a search is to be made (e.g. title words, authors, control numbers).

If the search is to be made for a particular journal part, then it needs agreement on an unambiguous syntax to represent the holdings. There should be two possibilities equating to the 'compressed' and 'uncompressed' of USMARC for Holdings. Both of these should be treated as distinct elements as they will need separate processing to retrieve and then match them against the searcher's requirements.

The SICI has been added to Approved extensions to bib-1 Attribute set (dated April 1997). This covers only the standard SICI which is equivalent to uncompressed. This is all that would be needed for a search element. However it would be useful to be able to retrieve via the MARC record a list of related SICI's like those relating to the journal part or adjacent journal parts.

4.3.2 MARC record returned

Discussion has taken place above regarding the requirements for the MARC record which Z39.50 requires to be returned.

When the MARC record is returned with its holdings information, as well as indicating for each distinct holdings field whether the contents of that field are displayed in compact form or otherwise, it would be ideal to indicate the kind of compliance they

have achieved as far as accuracy is concerned, and to what standards they conform. A subfield containing a code indicating level of compliance could include such data. This code could be added to each repeated field indicating the compliance for the record as a whole in case only one field for a particular serial item is included in the retrieved MARC record.

5 Conclusion

The SICI and MARC options described here should be seen as exclusive alternatives as it would be awkward to have to live with each.

Work should be done on the feasibility of converting existing 'eye-readable' records which are available as computer records to the formats prescribed.

It is important to remember that matching is easier against records which conform to a standard and that it is easier for each metadata source to convert its own records to a standard than for a search algorithm to interpret non-standard or semi-standard data in holdings statements. It is better to convert existing holdings statement to a standard format than to do nothing because they are non-standard. At the same time it is better for future holdings to be stored in a standard format where the data are present for individual issues.

In the feasibility and then the ease of matching lies the crux of success in the retrieval of serial parts. There is no other way than the provision of data which conforms to a certain standard in form and completeness. Attaining this level by upgrading existing records is the only way. The extent to which existing records can reach these proposed standards by algorithm and the extent to which manual intervention would be necessary is an important question which this study has not attempted to answer because the differing formats of metadata and their various levels of detail mean that each source has to work this out for itself.

References

1. *Specification for serials holdings statements for library and documentation centres*, London, BSI, 1980 (BS 5999)
2. ISSN. The ISSN (International Standard Serial Number) is maintained by the ISSN Centre in Paris. Information about ISSN is published in: *International Serials Data System. ISDS Manual*. Paris: ISDS International Centre, 1983. Details at <http://www.issn.org/broce.htm>
3. CODEN. The CODEN is an identifier similar in function though not in appearance to the ISSN which predated it. It is maintained by the International CODEN Service which is based at the Chemical Abstracts Service. <http://www.cas.org/coden>
4. ISBD. The International Standard Bibliographic Descriptions are cataloguing codes dealing with the description of documents. The main one on which the

others are modelled is *ISBD(G): General International Standard Bibliographic Description*. Rev. ed. Munich, K.G. Saur, 1991

5. Unesco. *UNISIST Reference Manual for Machine-readable Bibliographic Descriptions*, compiled by M. D. Martin. Paris: Unesco, 1974. Third edition, compiled and edited by H. Dierickx and A. Hopkinson. Paris: Unesco, 1986.
6. CCF is defined in: *CCF/B: the Common Communication Format for Bibliographic Information*. Paris, UNESCO, 1992
7. ASIDIC/EUSIDIC/ICSU-AB/NFAIS. *Recommended Interchange Specifications for Computer Readable Bibliographic Data Bases*. April 1978.
8. UNIMARC is defined in: International Federation of Library Associations and Institutions. *UNIMARC Manual*. London, IFLA UBCIM Programme, 1987.
9. UK MARC. The UK MARC format maintained by the British Library and defined in *UK MARC manual*, 3rd ed. London, British Library, 1990
10. USMARC. USMARC is defined in a number of documents of which two concern us here: *USMARC format for bibliographic data*. Washington, DC, Library of Congress, 1994 (updated 1996) and a holdings format (see reference 22).
11. *Anglo-American Cataloguing Rules*. London, Library Association, 1967
12. *Anglo-American Cataloguing Rules*. 2nd ed. London, Library Association, 1978
13. *London Union List of Serials*. telnet://uls.qmw.ac.uk
14. *Information and documentation: Holdings statements: summary level*. Geneva, ISO, 1997 (ISO 10324:1997)
15. *Serial item and Contribution Identifier. Version 2: ballot draft*. Washington, ANSI/NISO, 1997. (ANSI/NISO Z39.56-199X). Available on Internet at <http://sunsite.Berkeley.EDU/SICI>
16. *Data elements and interchange formats: information interchange: representation of dates and times*. Geneva, ISO, 1988 (ISO 8601:1988)
17. *Information retrieval (Z39.50): application service definition and protocol specification*. Washington DC, Z39.50 Maintenance Agency, July 1995. Official text for Z39.50-1995
18. *Approved extensions to bib-1 Attribute Set*. Washington, DC, Z39.50 Maintenance Agency, 1995

- 19 International Organization for Standardization. *Format for Bibliographic Information Interchange on Magnetic Tape*. Geneva, ISO, 1981 (ISO 2709-1981)
20. BLCMP MARC is the version of UK MARC used by the BLCMP system
- 21 OCLC MARC. A variant of USMARC used by members of the OCLC Cataloguing Cooperative.
- 22 *USMARC Format for Holdings Data including guide-lines for content designation*, prepared by the Network Development and MARC Standards Office. Update 2. Washington, DC, Library of Congress, 1994. Looseleaf
23. *Serials holdings statements*. Washington DC, ANSI, 1986. (ANSI Z39.44)
24. *Holdings statements for non-serial items*. Washington DC, ANSI, 1989 (ANSI/NISO Z39.57-1989)
25. *Serial holdings statements at the summary level*. Washington, DC, ANSI, 1974. (ANSI Z39.42). Now withdrawn
26. *British Library, Library of Congress and National Library of Canada achieve major milestones in format harmonization*.
<http://www.bl.uk/services/bsds/nbs/marc/marcpres.html>
27. *Information and documentation: bibliographic description and references: rules for the abbreviation of bibliographic terms*. Geneva, ISO, 1994
28. *Dublin core metadata element set home page* at
http://www.oclc.org:5046/research/dublin_core/
29. *Information processing: text and office systems: Standard Generalized Markup Language (SGML)*. Geneva, ISO, 1986 (ISO 8879:1986)
30. *MARC DTDs: Document Type definitions: background and development*. Library of Congress Network Development and MARC Standards Office.
<http://www.loc.gov/marcdtd/marcdtdback.html>
31. Pira International are working with the UK Book industry to define SGML standards for information exchange. <http://newmedia.pira.co.uk/stan.html>
32. Murray, R and Davidson, E. *MODELS Library Interoperability Profile Family*. Fretwell-Downing, 3 April 1997. Section 5.
33. McCallum, S. Personal communication, 17 April 1997

34. Cathro, Warwick. Personal discussion at TC 46 meeting in London, May 1997.
National Library of Australia where USMARC has been adopted have not made use of the Holdings format.

