

Information Development

<http://idv.sagepub.com/>

Developing the Common Communication Format

Alan Hopkinson

Information Development 1986 2: 99

DOI: 10.1177/0266666698600200207

The online version of this article can be found at:

<http://idv.sagepub.com/content/2/2/99>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Information Development* can be found at:

Email Alerts: <http://idv.sagepub.com/cgi/alerts>

Subscriptions: <http://idv.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://idv.sagepub.com/content/2/2/99.refs.html>

DEVELOPING THE COMMON COMMUNICATION FORMAT

Alan Hopkinson

The purpose of a bibliographic exchange format is to promote the free flow of information by facilitating the exchange of bibliographic records in machine-readable form. There are already numerous exchange formats, both national and international and for specific subject areas and specific types of institution. Why, then, do we need another? This is a criticism which has, from time to time, been levelled at the Common Communication Format, a format which has been developed by a group of experts convened by Unesco and working in the context of its General Information Programme.

Most formats cater really well for only half the information community — either for libraries, on the one hand, or for secondary services (abstracting and indexing services) on the other. The Unesco Common Communication Format (1) is an exchange format intended for use by agencies that have records of monographs and serials (library-type materials) as well as records of journal articles, contributions in proceedings and other parts of physical documents which constitute intellectually-independent entities; i.e. the kinds of records created by abstracting and indexing services.

It is intended to enable the exchange of records between different agencies throughout the world and between developed and developing countries.

HISTORY AND DEVELOPMENT

Unesco was particularly interested in formats for the exchange of bibliographic data because it had received many requests from agencies around the world for advice on setting up national or regional bibliographies. When recommending which format such systems should use, consultants were encountering difficulties since formats were divided into two categories, those used by national libraries, the MARC family of formats, and those used by abstracting and indexing services which, though more diverse, were to a certain extent exemplified by the UNISIST *Reference Manual* format. (2) This format had been developed jointly by Unesco and the International Council of Scientific Unions Abstracting Board (ICSU-AB) to meet the need for standards and guidelines for secondary services which were considering the automation of their databases.

In 1978, Unesco joined with ICSU-AB, the International Federation of Library Associations and Institutions (IFLA) and the International Organization for Standardization (ISO) in sponsoring an

International Symposium on Bibliographic Exchange Formats to consider the problem of the existence of the two categories of format which reflected the two different information communities, libraries and secondary services. (3) It was clear that in the future there had to be more cooperation in the standardization of bibliographic records if the needs of the end-user were to be satisfied. An end-user could be exemplified as a research worker wanting articles on a particular subject and then the serials in which they were to be found. Researchers tend to find references to articles in the databases of abstracting and indexing services, but have to go to a library to find the appropriate journal, conference proceedings or collection of essays, etc.

The Symposium resolved to attempt to break down the barriers between the two information communities. One way to contribute to this was by devising a format which would not be directed specifically to either community and which could therefore be used by any agency which was providing records to both. The *Ad hoc* Group on the Establishment of a Common Communication Format was therefore convened. The Group compiled a data element directory which took into account the various international exchange formats including UNIMARC (4), the USSR/US Common Communication Format, MEKOF (5) (the format of the eastern European countries), the format of the International Serials Data System (6), and the UNISIST *Reference Manual*. A KWOC index to the data elements was prepared and an abbreviated definition for each data element as found in each format was included. These became the basis for discussions which revolved around the definition of a mandatory core set of data elements. Optional data elements were then added to make it possible for the format to carry complete bibliographic records.

It quickly became clear that for the format to be accepted it would not be possible to prescribe exactly the form and content of every data element. Although it was possible to be precise in the case of some elements, such as those whose form and content were already prescribed in other standards (e.g. International Standard Book Number-ISBN, International Standard Serial Number-ISSN and key title), consensus could not be achieved in respect of others. In fact, this consensus could never be achieved without there being universally-accepted cataloguing rules. When the Common Communication Format was published by Unesco in 1984, the representation of many of the data

Developing the Common Communication Format

elements was therefore specified as being 'in accordance with the practices of the agency preparing the record'. For this reason, an agency receiving records in the CCF format would need to know the rules used by the source agency in order to determine whether it was worth taking the records.

For shared cataloguing, it is important that all records should conform to a particular cataloguing code, but there are other uses for bibliographic records where this is not so important. If records are in one and the same format, and therefore have the same identifiers, they can easily be interpreted and printed out on the computers of the agency receiving the record. If the records originate in a library, they can then be used to establish what that library has in its book stock; in the case of records from a secondary service, they can be used to produce current contents listings of journal articles. They can also be added to on-line databases, though the absence of universally-accepted rules for access points (or headings as they have traditionally been called in cataloguing) may make information retrieval a little more tedious than it otherwise would be.

It is only when records from different sources and prepared according to different cataloguing rules are merged into a database with a view to producing a printed catalogue that problems arise. Records created under different rules cannot easily be filed together in a printed product and duplicate records for the same document may get into the database

because they look different and avoid the checks made to detect duplicates. Standard numbers (ISBN and ISSN) are useful in detecting duplicates, but much grey literature has no standard number, not to mention the material published in those countries that have not yet or only recently set up an ISBN Agency. In any case, standard numbers did not exist at all before the late 1960s and so are never found on older material. The International Standard Bibliographic Descriptions ensure uniformity only in the descriptive area of the record and are, in general, only used by the library community. (7) These go some way towards ensuring that records from different sources which relate to the same document will be recognizable as such, and are of particular value since they have been incorporated into many national cataloguing codes. However, there have been no advances towards universal standards in the form of headings. The Common Communication Format, developed in the present situation where there is little standardization of record contents, has therefore been intentionally flexible in this respect.

IMPLEMENTING THE CCF

Exchange formats do not usually specify how data are to be input and the CCF is no exception. Implementation manuals, therefore, are needed to show how to create bibliographic records for entry into the format. Systems which do not already have

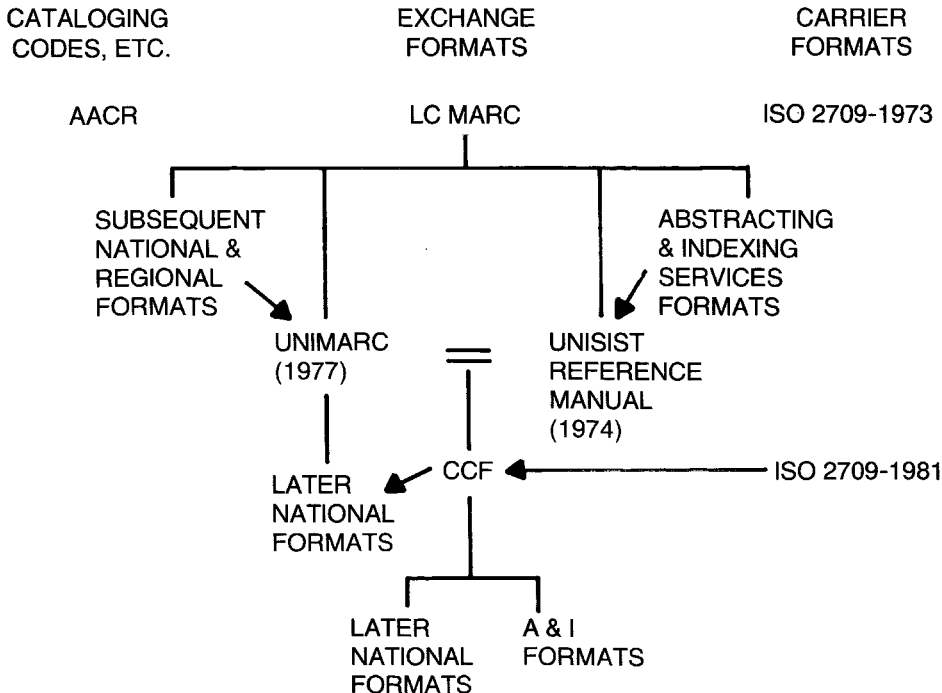


Figure 1. Interrelationships between formats. (13)

Elements of an exchange format

Exchange formats consist of three components:

- (a) Rules for the arrangement on a computer medium of data to be exchanged (including specifications for the physical medium, which may be paper tape, magnetic tape, disk, diskette, or even a transitory medium as in the case of on-line exchange).
- (b) Codes to identify the different data elements in the record (e.g. author, title, scale of map, starting date of journal, etc.).
- (c) Rules for the formulation of the different data elements — very closely tied up with (b). The data elements separately identified by the codes in the exchange format have to be defined, not only in terms of content but also in form if the records are to be suitable for use by another agency.

Effective exchange of bibliographic data between agencies can be accomplished only if the records of the agencies conform in respect of all three components.

In the case of the first, known as the carrier format, there exists a standard format for the exchange of data on magnetic tape which has been established by the International Organization for Standardization, ISO 2709. The universal acceptance of this standard has benefited the

information community enormously. It is accepted for the exchange of bibliographic data on magnetic tape, and it is also being used, as far as it applies, for the formatting of bibliographic data sent on-line.

The second component consists of the tags, indicators and sub-field codes — in short, codes which define the different data elements in the record. There is no universally-accepted standard for these. One reason why they vary between different implementations is said to be that agencies have different requirements in respect of the type and form of the data they wish to exchange. The MARC format was originally designed by the Library of Congress to enable the exchange of records between the Library and other institutions both in the USA and in other countries. There are also formats developed, or specifically used, by abstracting and indexing services like the Chemical Abstracts Tape Format. They all embody different schemes of tags and other identifiers.

The last component, the form and content of the data elements, varies according to the cataloguing rules used as well as according to the way the different data elements prescribed by the rules are divided up and separately identified by the format.

records in machine-readable form may use such manuals to help them create good quality records in the Common Communication Format, whilst those which already have machine-readable records may use them to convert their records into a form which will be compatible with CCF records from other sources. The CCF document which was published in 1984 is not intended to be a manual for data entry, nor to indicate how data should be converted into the CCF. It is more like a standard giving the bare bones of the Format. The order of the tags in the CCF is not based on the order in which a cataloguer using one of the major cataloguing codes would enter data. The development of the CCF concentrated on making the format compatible with other exchange formats, perhaps at the expense of making the format as it stands less suitable for data entry.

The CCF also includes a sophisticated record-linking technique requiring complex codes which should be entered by the computer rather than the cataloguer. (In the same way, no cataloguer is ever given the task of entering the codes in the MARC record directory which indicate the length of each data field and the character position within the record at which that data field begins). The CCF record structure was devised to enable record linking: this involves the analysis of records into their different

bibliographic levels to a much greater depth than in traditional cataloguing systems. Abstracting and indexing services, however, have developed systems which make full use of bibliographic levels. Both the International Nuclear Information System (INIS) of the International Atomic Energy Agency (IAEA) and the International Information System for the Agricultural Sciences and Technology (AGRIS) of the Food and Agriculture Organization of the United Nations (FAO), for example, employ formats and systems which require a records to be divided into its bibliographic levels. However, the CCF implements bibliographic levels in a way that has only recently become possible within the context of ISO 2709, the standard for bibliographic data interchange. This is because an extension to the record directory incorporated into the second edition of the standard (ISO 2709-1981) (8) enables pointers to group together parts of the record into what the CCF calls 'segments'. These can be used to separate data relating to the serial and the monograph when a record relates to a monograph in a serial, or data relating to a serial and its earlier and later titles when the record records all data about the bibliographic history of a serial. The fourth digit (immediately after the three-digit tag) clearly identifies which fields relate to the monograph and which to the articles: in the

Developing the Common Communication Format

CCF, when records are in exchange format, this code, known as the segment identifier, is found in the record directory, immediately after the tag, followed by pointers which 'point to' the location of the associated data. These pointers are necessary since records in the exchange format are arranged so that all the tags are placed together near the beginning of the record and all the data fields, which will eventually form the text of an entry in a catalogue, are placed together after the tags.

This record structure has to be interpreted for the catalogue to be able to use it. So far, a number of users and intending users of the format have designed implementation manuals to specify how the CCF has to be used. The resulting documents enable records to be produced which are identical in specification to or at least compatible with (i.e. convertible to) the CCF. The Office for Official Publications (OP) of the European Communities, for example, has produced FORMEX, the purpose of which is to 'provide a detailed and structured method for recording information about the OP publications in a computer-readable bibliographic record, for exchange purposes between two or more computer-based systems'. (9) The format goes beyond the CCF in that it includes fields for full text of the document, and conventions have been devised to enable the full text to be printed out in exactly the same form as the original document. The UN Dag Hammarskjold Library in New York has prepared the

UNBIS *Reference Manual for Bibliographic Description* which enables cataloguers to catalogue records according to Anglo-American Cataloguing Rules and code them with CCF tags. (10) Because the software they use is not hospitable to the use of sub-fields, the manual either specifies a separate field for each CCF sub-field or defines sub-elements by the use of ISBD punctuation. Nevertheless, the records produced are compatible with the standard CCF record. A manual has also been produced specifying the International Construction Database (ICONDA) communication format for the exchange of records in the framework of ICONDA. (11) This manual is based on the CCF and FORMEX documents. The CCF proved attractive to the ICONDA group because of its flexibility and the fact that it did not specify that a particular cataloguing code should be used but instead left the detailed definitions of data elements to the users.

THE CCF AND OTHER FORMATS

Why do we need another exchange format when we already have UNIMARC and the UNISIST *Reference Manual*, not to mention other national and international exchange formats? To answer this question it is necessary to compare the other formats with the CCF.

UNIMARC is designed to serve national libraries with their own national formats who wish to

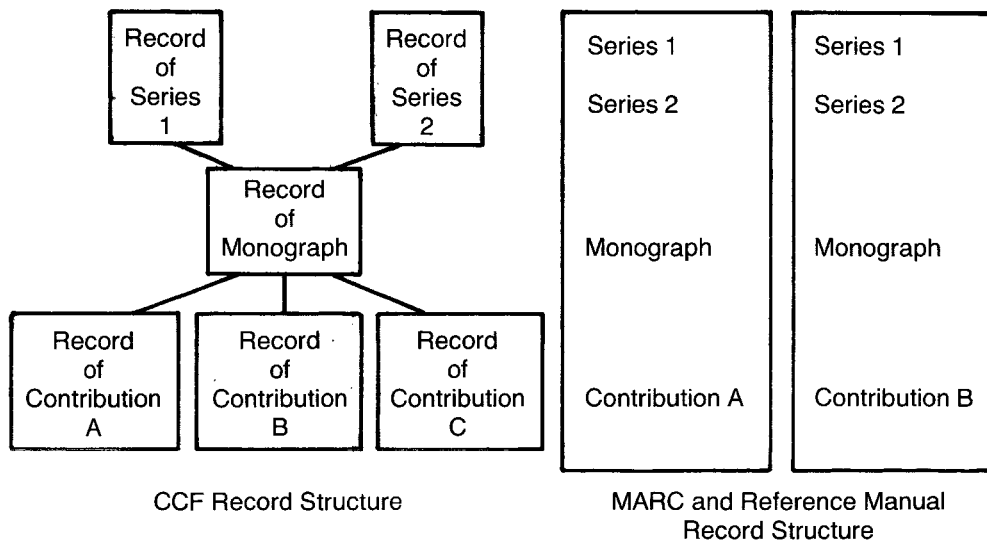


Figure 2. Comparison of record structures. (The records in the MARC and Reference Manual formats are repeated for Contribution C)

exchange records among themselves. Were it not for UNIMARC, exchange between national libraries would have to take place in every instance on a bilateral basis, with conversion programs being written between every pair of formats between which exchange was taking place. An alternative would be for every one to use one national format, for example, US MARC, since this was the first to be devised, but if that method were chosen there would be an international outcry every time the chosen national format was changed, since it would mean adjustments to everyone else's programs. UNIMARC is sufficiently close to most, if not all, national MARC formats to enable users to avoid complicated conversions between it and the national formats, and it can only be changed by agreement of the IFLA Sections on Cataloguing and Information Technology. The main disadvantage of UNIMARC is that it is biased, naturally, towards library materials and there is as yet no easy way to link a record for a periodical article to the record for the periodical which contains it. AACR-type analytics, which consist of an access point to a work published separately elsewhere, though here contained in a volume with other works, are no problem, but anything more complex like a set of articles in the issue of a journal becomes a little unwieldy.

The UNISIST *Reference Manual* specifically addresses the problem of recording articles which are found in journals or collections, but finds the solution in the only way possible at the time the *Manual* was compiled, when tapes were not only the sole medium for exchange but were also much used for textual processing in situations where disks are used today. In short, each article has its own bibliographic record which includes an abbreviated record of the item containing it. In a database organized in this way, there is a great deal of repetition which increases the likelihood of inconsistency caused by the possibility of the repetitions containing errors.

The CCF, on the other hand, has a mechanism to take care of record-linking in such a way as to avoid repetition. It can link together records at separate bibliographic levels which, when added together, constitute the record of a bibliographic item; or it can link records with different kinds of relationships such as a series title to its former title and *vice versa* or a work to its translations. The linking mechanism has been devised to be flexible in order to make it compatible with the complex mechanism of UNIMARC and the 'flat' record structure of the Reference Manual, which assigns a different tag to the same data element depending on whether it is at the monographic, analytic, serial or collection level. In addition, the CCF had to take into account the record structure of MEKOF.

The flexibility with which the CCF is endowed means that the user has to be careful. It is essential to ensure that the linking mechanism is used in a consistent way or else recipients of the records may

not be able to build into their programs everything necessary to decode the particular mechanisms being used. Since the records are split into 'segments' which point to other records, it is necessary to ensure that all the records which make up a complete bibliographic record are exchanged together, though it is possible to use the CCF without referring to any other records from within a record. More information and examples on this linking mechanism are found in the CCF document itself.

Although the linking mechanism is the outstanding feature of the CCF, it has another feature which has proved attractive to some potential users: the fact that it is truly independent of cataloguing rules. A network which already has its own rules for bibliographic description will find it easier to adopt the data element definition of the fields in the CCF than those of other exchange formats. The CCF here contrasts markedly with UNIMARC. As mentioned above, for the production of printed catalogues, all records should ideally be created according to one cataloguing code. This is clearly not possible in international exchange. UNIMARC needs to overcome such problems since its aim is to be an exchange format for national library cataloguing databases. When the *UNIMARC Handbook* was being prepared, (12) attempts were made to overcome these problems by giving instructions to make the data elements as little as possible cataloguing-code based. But this was only successful inasmuch as the cataloguing codes of the different major national libraries are not dissimilar (this has been especially true since they incorporated the ISBDs). However, it means that UNIMARC is really suitable only as an exchange format for good-quality records prepared according to national cataloguing rules.

Consequently, there is still room for all the universal formats which already exist; and formats such as AGRIS and INIS which serve international subject-based networks should certainly continue to be used.

THE FUTURE

The CCF was published in mid-1984 and already a number of organizations are planning to use it — a testimony to its effectiveness considering that the choice and adoption of standards is usually a process that takes years rather than months. In addition to the systems developing implementation manuals mentioned above, an exercise is under way in Brazil to update the national CALCO format (a MARC-type format used by library and secondary services alike) and bring it into line with the CCF.

The ultimate aim of an exchange format, as noted at the beginning of this paper, is to promote the free flow of information. It seems that the CCF is likely to make a significant contribution to this as a bridge

Developing the Common Communication Format

between different practices found in the library and information communities.

References

1. *CCF: the Common Communication Format*. Paris, Unesco, 1984 (PGI-84/WS/4).
2. Martin M.D. *Reference manual for machine-readable bibliographic descriptions*. Paris, Unesco, 1974. 2nd edition, edited by H. Dierickx and A. Hopkinson. Paris, Unesco, 1981 (PGI-84/WS/20).
3. International Symposium on Bibliographic Exchange Formats. *Towards a common bibliographic exchange format?: proceedings*. Budapest, OMKDK, 1978.
4. *UNIMARC: Universal MARC Format*. London, IFLA International Office for UBC, 1977.
5. International Interchange Format - MEKOF-2: specification of data elements. Moscow, ICSTI, 1977 (Official English translation of *Mezhdunarodnij Kommunikativnij Format*).
6. *ISDS Manual*. Paris, Unesco, 1984.
7. *Six International Standard Bibliographic Descriptions* have been published by the IFLA International Office for UBC, covering monographic publications, serials, non-book materials, cartographic materials, antiquarian materials and printed music. There is also an ISBD(G), a General framework which governs the structures of the other ISBDs.
8. International Organization for Standardization. *Documentation: format for bibliographic information interchange on magnetic tape*. 2nd edition. Geneva, ISO, 1981 (ISO 2709-1981).
9. *FORMEX: formalized exchange of electronic publications: working document*. Luxembourg, Office for Official Publications of the European Communities, 1985.
10. UNBIS. *Reference manual for bibliographic description: a manual for the preparation of bibliographic data for input into and retrieval from the United Nations Bibliographic Information System*. New York, UN Dag Hammarskjold Library, 1985.
11. ICONDA. *Communication format: format for the exchange of records in the frame of the International Construction Databases*. First edition draft. Stuttgart, IRB Verlag, 1985.
12. Hopkinson, Alan. *UNIMARC Handbook*. London, IFLA International Office for UBC, 1983.
13. The 'family tree' in Figure 1 shows the relationship between different bibliographic standards which are related to formats for the exchange of bibliographic data. It shows the direct descendents of MARC which was

developed by the Library of Congress in the USA in 1967. This culminates in the 'marriage' of UNIMARC and the UNISIST Reference Manual resulting in the production of the Common Communication Format. The success of the MARC experiment was such that it spawned formats in both the national library and the abstracting and indexing communities. Unlike human kinship, formats have been subject to cross-fertilization. The arrows in the diagram indicate influence. Thus UNIMARC, though descended from LC MARC, has been influenced by national and regional formats. ISBDs have also influenced UNIMARC. Interestingly, the earlier national formats were based very closely on LC MARC, later ones on UNIMARC, and more recent ones have been influenced by the CCF, though in the case of Brazil and Colombia they were originally based very closely on LC MARC and were in suspended animation for a number of years. The standard record structure, ISO 2709, was based on the record structure of the LC MARC format. The second edition has an optional, more complex structure, which has been used in — and has indeed made possible — the development of the CCF within the context of international standardization.

Abstract

The Unesco Common Communication Format (CCF) is described in the context of other exchange formats. A definition is given of 'exchange format', and the CCF is compared against this definition. The history of its development is outlined and its major technical features are summarized. Examples are given of the ways in which it is being used and is likely to be used in the future, and a number of implementation manuals are mentioned which have been developed to assist in its use.

Alan Hopkinson first became involved in bibliographic exchange formats when he worked on the British Library MERLIN project. He developed that interest further while working in the UNISIST Centre for Bibliographic Descriptions (UNIBID) and subsequently as compiler of IFLA's *UNIMARC Handbook*. He has undertaken a number of consultancies for Unesco and other international organizations, and is now Information Systems Manager at the Institute of Development Studies, University of Sussex, England.