THE UNESCO COMMON COMMUNICATION FORMAT

ALAN HOPKINSON
British Library Bibliographic Services Division
Store Street, London WC 1C 7DG
United Kingdom

Unesco's Common Communication Format (CCF) like other exchange formats enables free exchange of bibliographical records within information community. To make it more effective, efforts should be for making conversions in existing formats and their easy availability. Studies the CCF in relation to other formats like UK MARC, AGRIS, INIS and MEKOF-2.

INTRODUCTION

The Unesco Common Communication Format (CCF)[1] is an exchange format intended for use by agencies within the information community which wish to exchange bibliographic records with one another. It is intended particularly for those agencies which 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; these records tend to be created by abstracting and indexing services.

An exchange format consists of three components:

(a) Rules for the arrangement on a computer storage medium of data to be exchanged (including rules for the size of the physical storage medium).

(b) Codes to identify the different data elements in the record (e.g. author, title, scale of map, starting date of journal).

(c) Rules for the formation of different data elements, very closely tied up with (b). The data elements separately identified by the code 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 among agencies can be accomplished only if the records of the agencies conform to (a), (b) and (c).

In respect of the first component, 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[2]. Universal acceptance of this standard has helped the information community enormously. It is accepted for the exchange of data on magnetic tape, and it is also being used for the formatting of bibliographic data sent on-line down the telephone.

The second component relates to the tags, indicators and subfield 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 among different implementations is that agencies have different requirements in respect of the data they wish to exchange. Most countries, for example, have found it necessary to develop their own national MARC formats, and there are also formats developed by or specifically intended for abstracting and indexing services. 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.

The Common Communication Format implements ISO 2709 in full. It has its own system of tags and other identifiers for the identification of data elements, but it does not assume any particular cataloguing code. This point will be elaborated later.
HOPKINSON

HISTORY AND DEVELOPMENT

The Common Communication Format has been developed by a group of experts convened by Unesco and working in the context of the Unesco General Information Programme. In 1978, the International Symposium on Bibliographic Exchange Formats[3], sponsored by Unesco took place. Unesco was interested in this topic, because they had received a large number of requests from agencies around the world for advice on setting up national or regional bibliographies. When recommending which format a system should use, experts were encountering difficulties, since format were divided, roughly speaking, 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[4]. This format had been developed jointly by Unesco and the International Council of Scientific Unions Abstracting Board (ICSU-AB), because the need was felt for standards and guidelines for secondary services considering the automation of their databases. The international symposium addressed itself to the problem of the existence of the two categories of format which reflected the two different information communities, libraries and secondary services. It was clear that in the future, at the level of standardization of bibliographic records, there had to be more cooperation if the needs of the end user, for example, a research worker wanting articles on a particular subject, and then the serials in which they were to be found, were to be satisfied. Researchers find references to articles in the databases of abstracting and indexing services, but have to go to a library to find the appropriate journal or monograph.

The symposium resolved to attempt to break down the barriers between the two information communities. One way to contribute towards this was the devising of a format which would be directed specifically to neither community and could be used by any agency which was providing records to both the communities. The Ad-hoc Group on the Establishment of a Common Communication Format was convened. The members firstly requested that Unesco commission of data element directory taking into account the data elements of various international exchange formats, including UNIMARC[5], USSR/US Common Communication Format, MEKOF (the format of the CMEA countries[6], the format of the International Serials Data System[7], and the Reference Manual. A consultant prepared a KWOC index to the data elements in the formats and an abbreviated data element definition for each data element. These became the basis for the discussions which revolved around the definition of a mandatory core set of data elements. Optional data elements were then added to make possible the carrying of complete bibliographic records in the format. However, it quickly became clear that for the format to be accepted it would not be possible to prescribe exactly the form and content of each element although it was possible to be precise in some data elements such as those whose form and content were prescribed by other standards, such as the standard numbers (ISBN and ISSN) and key title. Sufficient consensus could not be obtained in the case of other data elements. That could never be achieved without internationally accepted cataloguing rules, which are still something for the future.

In 1984, when the Common Communication Format was published by Unesco in the middle of the year, the representation of many of the data elements was specified as being 'in accordance with the particles of the agency preparing the record', which means each agency according to its own cataloguing rules. Thus, an agency receiving a record in the CCF would need to know the rules used by the source agency in order to determine whether it was worth taking these records. That is to say that records created according to different rules would be of no use at all to a receiving agency. Being in the same Common Communication Format and therefore having the same identifiers, records can easily be interpreted by the computer programs of a receiving agency. They can be used to find out what an organization has in terms of its book stock, assuming that the records originate from a library, or they can be used to produce current contents lists of journals, for example, if the records originate from a secondary information service. They can be
added to on-line databases, though the absence of common rules for access points may make information retrieval a little more difficult and less efficient. It is when such records from different sources, prepared according to different cataloguing codes are merged into a database with a view to producing a printed product such as a catalogue that problems arise. Records created under different rules cannot be easily filed together in a printed product. There do exist the International Standard Bibliographic Descriptions[8] which ensure uniformity in the descriptive area of the record, but we are still looking forward to the day when there are universal standards for the form of headings.

THE CCF AND SOFTWARE

The Common Communication Format is not of course an end in itself. Unesco was interested in its establishment as a means to an end. The format has to enable the exchange of records among different agencies throughout the world and between developed and developing countries. To be useful, what is required is a pool of records which can be drawn on by agencies to avoid their having to catalogue of their own records themselves, working on the same principles as the cooperative cataloguing projects found in some countries. For this to be effective, it needs a number of agencies, all provided with compatible software for data entry and information retrieval, and using a common format, at least for exchange. In an ideal situation, they would use the same cataloguing rules, but as yet this is not feasible. The Institut für Maschinelle Dokumentation in Graz, Austria, is developing a software package known as IV+V. Unesco is interested in this software package, since it can be applied to bibliographic information retrieval and has offered advice on the requirements to make it suitable for use in developing countries for organizations wishing to exchange records with one another. Therefore, one of these requirements is the facility to provide data output on a computer-readable medium formatted to the provisions of international exchange formats. Exchange formats do not usually specify how data are to be input and the CCF is no exception. Therefore, an implementation manual is needed to enable the creation of bibliographic records for entry into the system. This manual has to make data entry as easy as possible, which is very important if it is to be used by cataloguers far removed from the help of fellow professionals. Most importantly, the data entered must be capable of being transformed into the output as prescribed by the exchange format, for it to be useful to recipient organizations. At one extreme, data could be entered in the form exactly as required by the exchange format. This would cause difficulties in data entry. As mentioned above, the Common Communication Format is independent of any cataloguing code. To a cataloguer, however, a cataloguing code is a very important element in record creation, and so the format must be related to the code. Therefore, an implementation manual aimed at cataloguers has to be related as closely to the cataloguing rules as it is to the exchange format. UK MARC [9], for example, contains against each data element the relevant AACR[10] section number. The first implementation manual using the CCF will take AACR for its cataloguing rules, since AACR is one of the most widely used codes. However, AACR does not address all the conditions encountered especially by secondary services (Chapter 13 is not very detailed and is very much in the library tradition). Therefore, extra guidance has to be given, so that the manual can be used to create records which are compatible with AACR, but which will be acceptable in abstracting and indexing publications.

DESIGN OF THE IMPLEMENTATION MANUAL

The implementation manual enabling the entry of data in a format compatible with the CCF has required a different approach from that taken in data entry manuals for systems using other formats.

One reason for this is that most exchange formats such as national MARC formats have been developed simultaneously with their input format, and with particular cataloguing rules in mind. Therefore, the format corresponds very closely to the requirements of the cata-
loguing rules. The same goes for the exchange formats of abstracting and indexing services like AGRIS[11] and INIS[12]. They too are closely related to the input format and the cataloguing rules used. Users of the CCF are not, however, using a format designed with any regard for their data entry procedures. This causes problems which can be overcome by the software being designed to convert the data. The manual must, therefore, be directed towards the cataloguing rules and the input procedures rather than the exchange format.

In practice, cataloguers do not enter bibliographic data in the form in which it appears on an exchange tape. They usually enter data on a worksheet next to the tag to which they relate. On the ISO 2709 exchange tape, tags are held apart from the data to which they relate. The tags are associated with their respective data not by location but by means of pointers, strings of numbers held in a part of the record called the directory[13]. Although these numbers are easily calculated by the computer, they are difficult for a human being to calculate. Therefore, they are never input by the cataloguer. Nevertheless, in many formats, the cataloguer finds the codes on the worksheet, the tags, in numerical order, because they, unlike the pointers in the directory, were designed with an eye on input.

The order of the tags in the CCF is not based on the order in which a cataloguer using AACR enters data. The development of the CCF concentrated on making the format compatible with other exchange formats, perhaps at the expense of making the format suitable for data entry as it stands. The CCF also includes a sophisticated record-linking technique which requires complex codes which the computer rather than the cataloguer should enter. Its record structure was devised to enable record-linking; this involves the analysis of records into their different bibliographic levels to a much greater extent than cataloguers would take the analysis in more traditional cataloguing systems. This kind of record structure has only recently become feasible within the context of ISO 2709, since an extension to the directory incorporated in the second edition enabled pointers to group together parts of records into what the CCF calls 'segments'. This greater conceptual gap between the exchange format and the cataloguer's practice has made the development of an implementation manual for data entry amount to the development of a new format. Incidentally, had it not been geared to just one cataloguing code, the resulting manual would have been unwieldy and cumbersome to use.

**THE CCF AND OTHER FORMATS**

Why do we need another exchange format when we already have UNIMARC and the RM, not to mention national exchange formats, is a not infrequent criticism levelled at the CCF. However, both these formats cater really well for only half the information community. UNIMARC is ideally suited to serve the national libraries with their own national formats which wish to exchange records among themselves. Were it not for UNIMARC, exchange among national libraries would have to take place in every instance on a bilateral basis, with conversion programs being written between every pair of formats among which exchange was taking place. An alternative would be for everyone to use one national format, for example, US MARC, but if that method were chosen, there would be an international outcry every time the national format chosen was changed, since it would mean adjustments to everyone else's programs. UNIMARC is sufficiently close to the majority of, if not to all, MARC formats to avoid complicated conversions between it and the national MARC formats, and it can be changed only by international agreement. The main disadvantage of UNIMARC is that it is biased, naturally, towards library material and there is no easy way to link the record of an article to the record of the serial containing it, for example. AACR-type analytics which consist usually of an access point for a work published elsewhere, even when it is contained in a volume with other works, are no problem, but anything more complex becomes a little unwieldy. The Reference Manual specifically addresses this problem, but finds the solution to it in a way which belongs to the days when tape were not only the sole medium for ex-
change but were also much used in textual data processing for storing data where disks would be used today. Each bibliographic level constituting an item which someone might wish to retrieve has its own record, and within the same record are all the details as to the monograph, collection or serial in which it is found. Thus, there will be an abbreviated serial record in every record of a contribution within that serial. The data elements in the total record are identified by tags which indicate the bibliographic level of that data element. Thus, the title of a monograph has a different tag from the title of a serial and the same goes for author, collation details, etc. In a database organized in this way, there would be a great deal of repetition; on an exchange tape it does not matter so much; indeed knowing that every record is complete makes processing easier.

The CCF on the other hand does have a mechanism to take care of record linking[14]. It can link 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 serial title to its former title and vice versa, or a work to its translations. The linking mechanism has been so devised as to be flexible. It had to be flexible to be able to be compatible with the complex mechanism of UNIMARC and the 'flat' record structure of the Reference Manual, which assigns a different tag to a data element, depending on whether it is at the analytic, monographic serial or collection level. Conversion between the CCF and MEKOF was also taken into account: although the record structure of MEKOF is 'flat', it uses a five-digit tag to enable one digit to indicate consistently the bibliographic level of the data element. However, the flexibility with which the CCF is endowed means that users have to beware: they must ensure that they themselves use the mechanism in a consistent way, or else recipients of the records may not have built into their programs the correct mechanisms to decode the particular mechanisms used. Moreover, the records are split into 'segments' which point to other records, so it is necessary to ensure that all the little bits of a record are exchanged together. Thus, housekeeping procedures have to be developed between parties to an exchange before exchange can take place and at the moment there are no guidelines on these.

Consequently, there is still room for all three international exchange formats which have been mentioned in this paragraph – MEKOF is regarded as regional rather than truly international; and there is no question that AGRIS and INIS should give up their formats which have been devised with a particular system in mind.

THE FUTURE

The important aim of all exchange formats is to facilitate the free flow of information, in this case bibliographic references. For this to be achieved, probably the most important thing is for computers to be able to talk to each other, which is something to be achieved by experts outside the library and information community. After that what is required are internationally-agreed forms of headings, or authority files. At the same time, more effort should be put into making conversion among existing exchange formats (and here the national formats are included) and making them readily available. Then the CCF would indeed be an effective bridge between the library community and the information community and would certainly gain popularity as a format on which to base an internal format for the databases of both library systems and secondary services.

REFERENCES


8. International Standard Bibliographic Descriptions, 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 general, a framework which governs the structure of the other ISBD's.


11. Most exchange format documentation contains at least one illustration of a record in ISO 2709 structure. See for example reference manual. op cit, p.4.27.


For background information to the other bibliographic exchange formats mentioned see:


ABBREVIATIONS


INIS - International Nuclear Information Systems of the International Atomic Energy Agency of the UN.