
Managing Machine-Readable Archives: Progress with Description and Exchange Standards.

by Michael Cook¹
Archival Description Project,
University of Liverpool

Background

Archivists have come somewhat belatedly to the idea that there should be formal standards for description and for the exchange of data about their materials. Observing progress made in these fields in North America, British archivists began work on constructing the necessary instruments in 1984. The Archival Description Project was set up at Liverpool University, supported by funds from the British Library Research and Development Department and the Society of Archivists. The Project team has produced two successive texts of a Manual Of Archival Description, affectionately known as MAD. The second edition, MAD2, published in 1990, was published by Gower, and has received a reasonable degree of trialing²

The archival community in Britain, however, finds itself in a difficulty as regards the formal adoption of a standard. There is a National Council on Archives, and a working party of this, chaired by Dr. Kitching (who is Secretary to the Royal Commission on Historical Manuscripts), has recommended the adoption of MAD2. In a rather similar way, the Society of Archivists has issued signals of approval, and has asked its Professional Methodology Panel to carry out tests and development work. These measures are somewhat short of a formal endorsement, but they do indicate acceptance at a practical level, and show that there is a will to continue developing the work.

The second edition of MAD contains rules for the description of a number of special formats, commonly found amongst archives. These are:

- title deeds (legal documents transferring land)
- letters and correspondence
- photographs
- cartographic archives
- architectural and engineering plans
- sound archives
- film and video archives
- machine-readable archives

This section of MAD2 must still be regarded as experimental, and it has not yet received adequate trialing. The principles on which the rules and guidelines are based, are coherent over the whole body of MAD2 and will be

discussed later in this paper.

The MAD2 special formats are intended for use in general archives repositories and services, not in specialist institutions. This important restriction should be emphasised.

The second set of archival description standards which should be mentioned are the international ones. The International Congress on Archives held in Montreal in September 1992, received the text of two new standards:

1. **Statement of Principles regarding archival description** (the Madrid Principles). Since this had been debated by the profession since 1991, this text was adopted.
2. **General International Standard Archival Description (ISAD(G))**. This was received as a draft for dissemination and discussion.

The first of these texts is now will be available. The second, (ISAD(G)), is not immediately available as it is in course of publication. It is intended that there should still be discussion of the topics presented, so that the process of maintenance and development may proceed. ISAD(G) itself has indeed not yet received formal adoption, but since it is in its second draft, and has received considerable discussion all over the world, it must be regarded as being near completion.

The other main standard applicable to archives, which ought to be mentioned here concerns data exchange. This is the MARC format, an archival application which was developed in the USA in 1984. It has become widely used in North America to allow archival descriptions to appear in the bibliographic databases, RLIN and OCLC. These databases are not widely available in Britain as yet, and the resistance of archivists to bringing in a library standard has been such that up to now MARC has been virtually unused for archives. There has indeed been little opportunity for it to be used. This situation appears to be changing, and a version of UKMARC in the archival format (AMC) is due to appear in 1993³

Turning now to the management and use of machine-readable archives, few British archivists have yet had

much experience. The ESRC Data Archive and the Edinburgh Data Library have been almost alone in the field in this country. The Public Record Office had ambitious plans to establish a Data Archive department during the mid 1980s. These have not progressed as might have been hoped. Recently there have been signs of life from this quarter, and we are given to understand that the PRO's Computer-Readable Data Archive will be established in 1995, with public access in 1997⁵.

It is important to make clear that there is a distinction between machine-readable files and datasets (which are the material administered by the ESRC Data Archive and other similar services) and machine-readable archives. The latter, like archives generally, are materials produced by, and forming part of the activity of, an organisation of some kind (such as a government). Archives of any sort are therefore unlikely to be one-time studies, or to have enough individual distinctness to allow them to be treated as discrete objects, comparable with books. Archives belong together in aggregations, which owe their character to the administrative system which produced them. Some of the consequences of this distinction are discussed further below.

The guiding characteristics of MAD2

The work both of the Archival Description Project and of the ICA's Ad Hoc Commission on Archival Description, has shown that certain basic principles underlie all description of archives. International agreement on this, at least as far as traditional records are concerned, is quite remarkable. The description of machine-readable archives, therefore, is likely to require attention to these principles, if only to test their applicability to new materials. The following section attempts to summarise what the basic rules are.

1. Levels of Arrangement and Description.

The idea that there are standard levels of arrangement is not new. The concept was first indicated in Europe at the start of the 20th century, then clarified in the USA⁶. It has been rediscovered and republished in different forms ever since. MAD2 restates the principle, but also extends it. A table of levels is given which looks at first sight like the hierarchical continuum characteristic of a classification scheme, and numbered like one:

0 Repository level: suitable for combined descriptions covering more than one repository.

1 Management levels assemblies of archival groups brought together on the basis of some common feature, for the convenience of the repository. E.g Official/non-official archives, ecclesiastical archives, private papers. Subordinate groupings may be numbered using decimals of 1.

2. Group or collection level (internationally **fonds**): the archives of distinct entities. Subgroups (functional divisions within the group) are numbered using decimals of 2.

3. Series (within Britain, termed class): physically related sets of archives. Subseries are given decimals of 3.

4. Items: the unit of physical handling (volume, file, box).

5. Pieces: indivisible components; documents. Levels 4 and 5 may be used interchangeably in some cases.

The interesting thing about this table is its universality. Yet it is unlike a general classification scheme because it is tied to observable external phenomena at three points:

Fonds (level 2) always relates to the total archival product of a distinct entity (organisation or individual);

Series (level 3) are always the physically and systematically related product of an administrative activity, sets that belong together because of the way they were created and used;

Items (level 4) are always the physical units of handling.

No level of arrangement is compulsory; though in the Madrid Principles it is stated that the level of the **fonds** is "the broadest unit of description"⁷. Therefore, provided that we accept that the three levels above must always be set to correspond to the appropriate physical entities, any or all of the levels of arrangement can be used, above the fonds, or below the item, as convenient.

There can be problems in identifying what should constitute a **fonds**. MAD2 advises that administrative or political levels of dependence should be disregarded. Thus an overall or umbrella organisation can be the origin of a fonds, but so can organisations which are administratively part of it. An extreme illustration would be that the Government of a country could be the source of a fonds (provided that it did actually produce records as such); but so could any of its Departments, or even lower subdivisions, sections etc. If any organisation is complete enough in itself to produce its own archives, it can originate a **fonds 1**

2. The multi-level rule

The multi-level rule in MAD2 states that archival descriptions should normally embrace more than one

level of arrangement. This is fully consistent with the multi-level rule laid down in ISAD(G), and in the Madrid Principles. However, MAD2 has a further elaboration of the principle, which has an important use in the context of finding aids. This is the concept of the 'macro' and 'micro' description.

These two terms do not relate to the specific levels of arrangement which are being described, but to the relationship between them. For example, finding aids frequently contain descriptions at fond, series and item levels. In these, the macro-micro relationship has a triple form:

Fonds description: a macro description governing:

Series description 1: a micro description in relation to the above, but a macro governing.

Item descriptions: micro descriptions of items in series 1, governed by the above.

Series description 2.... etc

In the MAD2 models, guidelines suggest that these relationships of dependence should be demonstrated to the user by the use of narrower margins, left and right; this assumes a hard-copy finding aid using standard pages. That is a common situation but not the only one. The important thing is that in any given case, the macro and micro descriptions may relate to any level of arrangement: fonds/item; management group/fonds; item/piece, etc. It is therefore a misconception to regard the macro description as peculiar to the 'higher' levels of arrangement, and the micro to the 'lower' ones.

Macro descriptions are written from a different standpoint than from micro descriptions. Their standpoint is the aggregate (whichever it is). Micro descriptions give information specific to each case. In the example above, the fonds description will give information relating to the fonds as a whole (probably including provenance information, but this is a separate issue); it also gives all information common to the series which follow, in order to avoid redundancy

The series descriptions which follow have a dual character. In so far as they are micro descriptions, they deal with each series one by one, giving specific information. Each serie description then operates as a macro for the items which follow. As macros they give information which relates to the series as a whole, and common data for the items. Finally, the items give data specific to each case.

This rule has been explained at some length because it

makes it immediately clear that, and why, standards originating in library practice are not suitable for archival applications.

3. The data elements table and its structure

Archival descriptions require data of two different kinds: information about the origin, background, context and provenance of the archive; and information about its content. Descriptions must therefore be essentially structured. The Project team drew up a list of the data elements that can be found in these descriptions, and drew them together into seven 'areas'. Like the levels, most data elements and areas are optional, and are brought into use only when required for the specific case.

MAD2 sets out a number of models which govern the way in which descriptions can be set out, using the data elements and areas. These models accommodate the multi-level rule and allow the dependence of micro upon macro descriptions to be demonstrated so as to be easily perceived by users.

4. Access Points and Provenance

ISAD(G) introduces the concept of access points, which should be subject to authority control. Access points should be provided for provenance information as well as for data from the contents of documents. Work on authority files, sadly lacking in the archive world, is therefore needed.

Standards for the description machine-readable archives

Unless it is true that machine-readable archives are quite unlike any other archive, description standards for them should follow the models and rules for archival description, including the basic principles outlined above. Section 25 of MAD2 deals with this problem.

Although it may be anomalous to speak of levels of arrangement where the materials can never be physically arranged, it is nevertheless true that there must be levels of description. Both the **fonds** (the archive of a whole organisation) and the series still appear to have a real existence.

There is some debate about whether or not machine-readable archives must be treated in a radically different way from other archives⁹. Those who concentrate on the media which carry electronic documents, are conscious above all of its evanescence, its lack of objective existence. Those who look primarily at the origin, context and purpose of the document will have a much more traditional picture. The **fonds** will doubtless also contain descriptions of traditional archives, or archives in alternative forms. The series is normally the dataset which can most be regarded as a complete entity for

description and management purposes. It most clearly resembles the datasets held by the ESRC Data Archive.

MAD2 proposes that there should be short descriptions of the entities at these two levels, written into the main finding aids of the repository. When this is done, separate and specialised descriptions of the machine-readable groups and classes can be established, with a linkage between the two systems. This method allows a generalist repository to have a finding aid system which is an effective intellectual control over its total holdings, while at the same time designing a specialist finding aid which is appropriate to technically different materials.

The specialist description may itself be multi-level, or it may be a flat file, according to circumstances. It must clearly contain all the metadata required: the technical information needed to record the internal structure of the file and its software dependence. The data elements needed for this are listed in Section 25.

A final note might be that background, context and provenance information should always be provided, because without it the meaning of the electronic record is lost. Indeed this point is conceded by the practice of Data Archives. This 'macro' information, however, does not necessarily have to be held in the detailed, specialised, file which is the direct finding aid to the machine-readable data. It may be held in the main finding aid system of the repository. In future, this main finding aid may of course be itself held in a machine-readable form; or it may be processed so as to enter it into a national index, or into a data entry system. For these, both cataloguing and data exchange standards will be needed.

1. Paper presented at IASSIST 93 in Edinburgh.
2. M.Cook & K.Grant. Manual of archival description. Society of Archivists, 1986. [Some exemplars of this edition were wrongly marked '2nd edition'] M.Cook & M.Procter. Manual of archival description 2nd ed. Gower, 1990.
3. Copies of texts and current drafts are obtainable from the Secretariat of the International Council on Archives Ad Hoc Commission on Archival Description, National Archives of Canada, Ottawa
4. Alan Hopkinson & M.Cook. Information from the former at the Library and Archive, Tate Gallery, London.
5. Alexandra Nicol and Steven Duffield. Unpublished paper to seminar on electronic records held at the School of Library Archives and Information Studies, University College, London, 10 Dec 1992.

6. Richard Lytle. **Subject retrieval in archives: a comparison of the provenance and content indexing methods**. PhD, University of Maryland, 1979.

7. Statement of principles regarding archival description, First version, revised, section 2.2.

8. Terry Cook. **Treatment of the archival fonds: theory, method and practice**. Bureau of Canadian Archivists, Ottawa, 2. It is interesting that the concept was known to, but misunderstood by, the Marxist regimes of Eastern Europe. They adopted the habit of setting the fonds at too high a level of institutional independence, hence most institutions had to be the originators of sub-fonds. This misuse serves to underline the validity of the concept when used properly.

9. Charles Dollar. New developments and the implication on information handling'. In Information handling in offices and archives, ed. Angelika Menne-Haritz. K.G.Saur, 1993 pp56-66.