

BIBLINK - LB 4034

D1.1 Metadata Formats

<i>Name of Client:</i>	European Commission
<i>Distribution List:</i>	Pat Manson, European Commission Project Partners
<i>Author:</i>	Rachel Heery with contributions from Robina Clayphan, Michael Day, Lorcan Dempsey, and David Martin.
<i>Authorised by:</i>	Ross Bourne, British Library
<i>Contractual Date:</i>	November 1996
<i>Date of Issue:</i>	23 December 1996
<i>Issue:</i>	1.0
<i>Reference:</i>	5003/DEL/56
<i>Workpackage:</i>	WP 1 - Study of Metadata
<i>Deliverable Number / Type / Nature</i>	D1.1 / P / Report
<i>Total Number of Pages:</i>	112
<i>Contact Details:</i>	Rachel Heery, UKOLN, University of Bath Claverton Down Bath BA2 7AY UNITED KINGDOM
	Telephone: +44 1225 826724
	Email: lisrmh@ukoln.ac.uk

TABLE OF CONTENTS

TABLE OF CONTENTS	i
1. Document Control.....	1
1.1 Abstract	1
1.2 Keywords.....	1
2. Management Overview.....	2
2.1 Executive Summary.....	2
2.2 Acknowledgements.....	2
2.3 Scope Statement.....	2
3. Objectives, Scope and Glossary.....	4
3.1 Scope and Assumptions.....	4
3.2 Work Available in Background.....	5
3.3 Glossary	6
4. Information Flow From Publishers to National Bibliographic Agencies.....	9
4.1 Organisations Involved in the Metadata Information Flow.....	9
4.1.1 Publishers	9
4.1.2 Information services.....	9
4.1.3 Suppliers.....	9
4.1.4 Libraries	10
4.2 Metadata Formats Associated with Particular Types of Organisation.....	10
4.2.1 Publishers, Booksellers and Suppliers.....	11
4.2.2 Libraries, National Bibliographic Agencies, Trade Bibliographic Agencies.....	11
4.2.3 Internet search services	11
4.2.4 Electronic Text Archives.....	11
4.2.5 Repositories.....	11
5. Typology of Metadata	12
5.1 Continuum of Complexity and Richness.....	12
5.2 Further Scoping Required.....	14
5.3 Topology of Metadata	14
6. Metadata Requirements of National Libraries.....	19
6.1 Responses to the Questionnaire	20
6.1.1 Cataloguing format and cataloguing rules used currently.....	20
7. Identification of Metadata Formats for Further Consideration.....	27

8. Description of Service and Technical Context for Metadata Formats.....	28
8.1 Simple Internet Descriptions.....	28
8.1.1 Dublin Core.....	28
8.1.2 IAFA templates.....	31
8.1.3 SOIF / RDM.....	33
8.2 Publishers Formats.....	35
8.2.1 MAJOUR SGML DTD for journal article headers.....	35
8.2.2 SSSH SGML DTD for article headers.....	35
8.2.3 BIC SGML DTD for non-serial publications ("books").....	37
8.2.4 EDIFACT EDI formats.....	39
8.2.5 Using EDIFACT to carry metadata.....	39
8.3 TEI, EAD and CIMI	40
9. Cataloguing Electronic Publications With MARC.....	41
9.1 Background.....	41
9.2 UKMARC.....	41
9.2.1 Those fields already accepted into the UKMARC manual:.....	41
9.2.2 Those fields already approved to be added to UK manual in an update:.....	42
9.2.3 Fields contained in British Library December 96 consultative document:.....	43
9.3 UNIMARC	44
9.3.1 Background	44
9.3.2 Fields used for cataloguing electronic resources.....	44
10. Mapping Metadata Formats to MARC	47
10.1 Mapping of IAFA Templates to OCLC MARC.....	47
10.2 Mapping of Dublin Core to USMARC.....	49
10.3 Mapping of Dublin Core to PICA Format.....	50
11. Conclusion.....	51
Annex A. Product Information Flows in the Publications Supply Chain.....	55
A.1 Background.....	55
A.2 Organisations involved in information transmission.....	55
A.3 Metadata information flow diagram.....	55
A.3.1 Books and other non-serial publications.....	57
A.3.2 New serials and changes to serials.....	58
A.3.3 Serial issue contents.....	58
A.4 Existing standards referred to in the text.....	59
A.5 Standards under development.....	59
Annex B. National Libraries Metadata Requirements: Questionnaire Replies.....	61
B.1 Bibliotheque nationale de France.....	61
B.2 Royal Library response.....	65
B.3 Norway response	75
B.4 Biblioteca Nacional.....	79
B.5 The British Library.....	81
Annex C. UKMARC Fields for Cataloguing Electronic Publications.....	84

1. DOCUMENT CONTROL

<i>Issue</i>	<i>Date of Issue</i>	<i>Comments</i>
v0.8	18 December 1996	Final version
1.0	23 December 1996	Formal authorised version.

1.1 Abstract

This report, written as a deliverable for Work Package 1 for the EC BIBLINK project, considers metadata referring to electronic resources in the context of the information flow from publishers to national bibliographic agencies. It makes recommendations to the project regarding metadata formats for use in this context. The report analyses the bibliographic link between publishers and national bibliographic agencies, in order to explicate current approaches to encoding metadata. It suggests a typology of metadata formats to assist in review of the characteristics of various metadata formats. The metadata requirements of national libraries are reviewed based on a questionnaire survey of partner libraries. Candidate metadata formats are identified for further consideration and the service and technical contexts in which they are used are reviewed. Possibilities for moving between formats are illustrated by various mappings.

1.2 Keywords

BIBLINK, metadata, library, bibliography, MARC, publisher.

2. MANAGEMENT OVERVIEW

2.1 Executive Summary

This is a summary of the recommendations made in this report:

Project partners need to consider these scoping issues and refine the scoping of the project to allow criteria to be drawn up for decision making.

Predicated on the stated assumptions as to scope, it is recommended that BIBLINK should concentrate on formats in Bands Two and Three for the exchange format. This does not preclude the possibility that conversion will be required from formats outside these Bands e.g. from more complex formats in Band Four into a simpler format, but that the formats for data exchange would be located in Band Two or Three.

As part of the consensus building process, publishers and national libraries should identify the objects and relationships which need to be represented in metadata describing electronic resources.

Project participants need to agree a clearer definition of:

- the level of detail required as regards content of metadata (simple CIP record or a richer record)
- what 'object level' national libraries will want to describe (the level of granularity of resource that will be described)
- to what extent the content of metadata can be covered by rules for formulation of content (i.e. cataloguing rules)
- reaffirm whether, within the project, the aim of the project is to catalogue electronic publications from a few well established publishers or to address improving bibliographic control of the wider world of web publishing.

We need to make explicit assumptions about the business model for the provision and exploitation of metadata.

It seems one metadata format may not be sufficient for the diverse body of publishers described in the scoping document. It would be more realistic to consider two formats to allow for the creation of a brief record and a more complex record.

Consider use of Dublin Core as a minimum element set. Consider use of BIC non-serial DTD and SSSH for more complex records. Consider implementation of Warwick Framework to package more complex SGML records with Dublin Core records.

We accept that all metadata formats in this area are unstable. We need to define what level of maturity and stability are required in our format(s) of choice. At that stage we may wish to influence the development of the format(s).

2.2 Acknowledgements

I should like to thank Brian Green (BIC) and Brian Holt (British Library) for providing background information for this report.

2.3 Scope Statement

This document is deliverable D1.1, Metadata Formats, for the BIBLINK project, and documents the findings of work package 1, 'Study of Metadata'. The objectives of this work package were to:

- identify the major current approaches to encoding metadata
- describe the service and technical constraints in which they are used
- perform a comparative analysis of their content
- make recommendations.

The report uses D0.1, 'Scoping Study', so focus the work in the areas of publication types and publisher types.

A study by the Text Encoding Initiative (TEI) to undertake a comparative analysis and produce a series of recommendations is also underway, and will be documented separately.

This report will have an impact on the following phase 1 work packages:

- WP3 - Consensus Building
- WP4 - Format Conversion Feasibility
- WP5 - Transmission of Data.

Compilation of this report has highlighted particular areas where refinement is required to the scoping of the project. These areas are indicated in the report particularly in section 6.1 and the Conclusion, and the various recommendations suggest where further consideration of scoping would be beneficial.

3. OBJECTIVES, SCOPE AND GLOSSARY

The objectives of this work package are to:

- identify the current approaches to encoding metadata
- describe the service and technical contexts in which they are used
- do a comparative analysis of their content
- make recommendations regarding metadata formats.

The outcome of this work package will feed into the consensus building process in WP3 which includes producing an agreed minimum set of data elements for the description of off-line and on-line electronic resources, and agreeing recommended formats for the exchange of data.

3.1 Scope and Assumptions

There has been a considerable amount of recent work and activity in the metadata area, particularly over the last year. It is not our intention to replicate existing documents, and wherever possible we refer to existing work. This report will consider the suitability of various formats but will not prescribe data elements at this stage. More detailed work on data elements is scheduled for WP3. In particular the final choice of which metadata format(s) will be used will be constrained by

- the requirements of national libraries and publishers
- the target set of publishers
- the target set of publications.

As part of this work package we have done some initial investigation of the requirements of national libraries, but as the project progresses these requirements will need to be defined and clarified. As yet we do not have a definition of the publishers' requirements as this will be built up during the consensus building process. For that reason, the suitability of various formats from the publishers' viewpoint will be suggested rather than matched against requirements. During the consensus building process it is anticipated that these criteria for choice of metadata will be refined to enable a final decision on metadata format. This work-package hopes to raise some of the issues involved in the choice, and provide background on the possible options.

Metadata is defined as "data which assists in the identification, description, evaluation and selection of an information object". Metadata can exist for objects at various levels of granularity e.g. it can refer to collections of documents, a single document or a chapter within a document; it can refer to a series, an individual journal, or an article within a journal; it could refer to a web-site, to a particular logical archive on that site, to a web page, or to an image embedded in that page.

At any given level of granularity, metadata may be required in a number of areas:

- intellectual content
- technical and physical characteristics
- access and availability data (e.g. 'host' server manager)
- terms and conditions
- evaluative data (e.g. PICS) etc.

In addition there may be metadata associated with the record itself. This includes such details as date record created, creator of record, etc. We call this data administrative metadata. No assumption is made that all metadata associated with an object needs to be contained within the same metadata record. Note that the use of particular identification schemes will not be considered as this is dealt with in WP2.

Different implementations may store and transport metadata records in different ways. For example, a record may be stored and transported within the object which it describes, or it may be stored and transported as a self-contained entity. Metadata referring to a single object may be subdivided into different packages, some of which are stored in separate locations and are referenced in the 'master' record. Decisions on the manner in which metadata is transferred will form part of WP5 (Transmission of metadata).

3.2 Work Available in Background

IFLA Section on Cataloguing. Functional requirements for bibliographic records: draft report for world-wide review. May 1996.

Models for the provision of national bibliographic services in Europe : final report / by Michele Lenart; translated and edited by the British Library. - Luxembourg : European Commission, 1996.

Metadata: a survey of current resource description formats. Work Package 3 of Telematics for Research project DESIRE (no. 1004). November 1996, at

<http://www.ukoln.ac.uk/metadata/DESIRE/overview/>

OCLC/NCSA Metadata Workshop Report. S. Weibel, J. Godby, E. Miller and R. Daniel, March 1995.

The March 1995 Metadata Workshop, sponsored by the On-line Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA), at

http://www.oclc.org:5047/oclc/research/publications/weibel/metadata/dublin_core_report.html

The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description.

L.Dempsey, and S.Weibel, D-Lib Magazine, July/August 1996, at

<http://www.ukoln.ac.uk/dlib/dlib/july96/07weibel.html>

A Syntax for Dublin Core Metadata - Recommendations from the Second Metadata Workshop.

L.Burnard, E.Miller, L.Quin, and C.Sperberg-McQueen,, April 1996, at

<http://www.uic.edu/~cmsmcq/tech/metadata.syntax.html>

BIC (Book Industry Communication) and British Library BNBRF Book Product Information project.

BIC's web pages provide background information, as well as their publications of standards and guidelines:

BIC Manual on Publishers' Bibliographic Databases, draft, 1994. (unpublished, but available on request from Book Industry Communication.)

Bide, Mark (on behalf of Book Industry Communication) Electronic Tables of Contents (EToCs) for serials: standards for structure & transmission. A research study for the BNB Research Fund. Fourth Draft 1994.

Martin, David. Data elements for an EDI 'Book Product Information' message. British National Bibliography Research fund Report 75. Published for BNB by BIC, 1995.

SSSH: Simplified SGML for Serials Headers, London, Book Industry Communication and PIRA International, 1996.

3.3 Glossary

This Glossary describes terms which are most relevant in the context of the BIBLINK project and this deliverable. Further information about the terms can be found in the following glossaries:

- A glossary of Internet terms by Internet Literacy Consultants (TM):
<http://WWW.matisse.net/files/glossary.html>.
- A glossary of computer oriented abbreviations and acronyms called BABEL by Irving Kind:
<http://WWW.access.digex.net/~ikind/babel96b.html>.
- A glossary for NCSA Mosaic and the WWW World Wide Web users:
<http://WWW.ncsa.uiuc.edu/SDG/Software/Mosaic/Glossary/GlossaryDL.html>.
- A glossary for the project Communication in Physics:
<http://WWW.phys.uva.nl/fnsis/onderzoek/comm/papers/glossary.htm>.

AHDS

Arts and Humanities Data Service.

Bibliographic description

a set of formalised data elements describing a publication.

Bibliographic record

a discrete bibliographic description stored either manually or electronically.

BIC

Book Industry Communications.

CIMI

Computer Interchange of Museum Information.

CIP

Cataloguing-In-Publication records, created using information supplied pre-publication by the publisher.

CD-ROM

Compact Disc Read Only Memory.

CURL

Consortium of University Research Libraries.

Database (DB)

a computer Program for entering, storing and retrieving items of information in a structured fashion.

Deposit of publications

a system in operation in most countries, usually legally enforced, whereby publishers must deposit one or more copies of every publication with nominated libraries. Often referred to as Legal Deposit.

DESIRE

Development of a European Service for Information and Research.

DTD

Document Type Definition.

EDI

Electronic Data Interchange. The exchange of structured data messages to enable automated transactions between application systems.

EDIFACT

EDI For Administrations, Commerce and Transport. The international EDI standard messaging syntax under the responsibility of the UN, for trading transactions in all industries. Also known as UN-EDIFACT.

Electronic journal

similar to a traditional journal but published only in electronic form - on a CD-ROM or the World Wide Web.

Electronic publisher

see publisher.

Electronic publication

document, file, journal, etc. made available in electronic form.

eLib

Electronic Library Programme, previously FIGIT (Follett Implementation Group on IT).

FGDC

The Federal Geographic Data Committee.

Format

in the context of bibliographic control, the formalised structure in which the specific elements of bibliographic description are accommodated.

FTP

File Transfer Protocol, an internet standard means of transferring electronic files between computers.

Home page

a World Wide Web page set up as an introductory page by an organisation or individual.

HTML

Hypertext Mark-up Language The standard language used for creating Web documents.

HTTP

HyperText Transfer Protocol. The protocol used for communication

IAFA	between Web clients and servers.
IEEE	Internet Anonymous FTP Archive.
IETF	Institute of Electrical and Electronics Engineers.
IFLA	Internet Engineering Taskforce.
ILL	International Federation of Library Associations and Institutions.
Internet	Inter-Library Loan.
ISAD	the world wide network of computer systems connected to each other.
ISBD	Information Society and Development (conference).
	International Standard Bibliographic Description. There are seven specific ISBDs as well as the general ISBD(G): monographs -(M), serial publications -(S), cartographic material -(CM), non-book material -(NBM), printed music -(PM), antiquarian publications -(A), computer files -(CF).
ISSN	International Standard Serial Number.
Legal Deposit	see Deposit of Publications.
MARC	MAchine Readable Cataloguing. A family of formats based on ISO 2709 for the exchange of bibliographic and other related information in machine readable form. For example USMARC and UNIMARC.
Metadata	information about a publication as opposed to the content of the publication; includes not only bibliographic description but also other relevant information such as its subject, price, conditions of use, etc.
Monograph	a publication either complete in one part or complete, or intended to be completed, in a finite number of separate parts. A non-serial publication.
Multimedia	a publication in which images, sound and text are integrated.
National Bibliography	a listing of all national publications. May include all publications produced in that country, or in the language of that country, or sometimes about that country.
NCSA	National Center for Supercomputing Applications.
NDIS	National Document Information Service.
OCLC	On-line Computer Library Center.
On-line resource	an on-line resource is an electronic document which is bibliographically identifiable, which is stored in machine readable form on an electronic storage medium and which is available on-line. For example - a Web page.
Off-line publication	an off-line publication is an electronic document which is bibliographically identifiable, which is stored in machine readable form on an electronic storage medium. For example a CD-ROM.
OPAC	Online Public Access Catalogue.
PICS	Platform for Internet Content Selection, an infrastructure for associating labels with Internet content.
PII	Publisher Item Identifier.
ps	postscript, a standard format for exchange of printable files.
Publications	documents containing either text or sound or images, or combinations of these, packaged for wider distribution, whether off-line (e.g. printed book, CD-ROM) or on-line (e.g. Web, database for information retrieval).
Publisher	a person or organisation that produces documents and makes them available. Newly emerging publishers may produce and distribute documents electronically - for instance, on the Web.

RDM	Resource Description Messages.
Record	see bibliographic record.
RFC	Request For Comments, a method by which standards (sic) are proposed and agreed, usually with reference to the Internet.
ROADS	Resource Organisation And Discovery in Subject based services
Serial	a publication in any medium issued in successive parts bearing numeric or chronological designations and intended to be continued indefinitely. Serials include periodicals; newspapers; annuals (reports, yearbooks, etc.); the journals, memoirs, proceedings, transactions etc. of societies; and numbered monographic series.
SGML (ISO 8879)	Standard Generalised Mark-up Language. ISO standard for document description, separating contents and structure.
SOIF	Summary of Object Interchange Formats.
SSSH	Simplified SGML for Serials Headers.
SURFnet	the Dutch national computer network for research and education in the Netherlands, joining local networks.
TEI	Text Encoding Initiative.
URL	Uniform Resource Locator. The standard way to give the address of a source of information on the WWW. It contains four different parts: the protocol type, the machine name, the directory path and the file name. For example: http://WWW2.echo.lu/libraries/en/libraries.html see Deposit of Publications.
Voluntary Deposit	
Web site	used to refer to a single location on the World Wide Web, usually on the same piece of hardware. Part of the Internet that stores and gives access to documents using HTTP.
World Wide Web	the global set of Internet web sites offering world wide access to information using HTTP.
Z39.50	A network protocol which allows searching of (usually remote) heterogeneous databases and retrieval of data, most often used for retrieving bibliographic records.

4. INFORMATION FLOW FROM PUBLISHERS TO NATIONAL BIBLIOGRAPHIC AGENCIES

Product information for hard copy material currently flows from publishers through various other agencies to libraries. It would be useful to identify the players involved in this process and identify the most widely used data standards and communication media. A detailed model has been drawn up (Annex 1) based on the present flow of information from traditional publishers for print and non-print publications in the UK. It is likely that the rise in popularity of electronic publication and the entry of 'new publishers' will cause some changes to this model, but the present model will remain, adapt and co-exist with other flows.

A more formal definition of 'publication' is required and will be explored in the context of BIBLINK. At this stage, we note a continuum between conventional print materials and newer forms of network publication which have characteristics unprecedented in other media. Physical carriers of digital material such as diskette and CD-ROM can be, and often are, distributed in exactly the same way and through exactly the same channels as printed books, audio and video cassettes, and other media - frequently as parts of a hybrid package.

This model is based almost entirely on the current situation in the UK, and there are probably some features which are at present unique to the UK. In the UK there is significant involvement of trade bibliographic agencies in the process of bibliographic control, this is not reflected, for example, in the Netherlands where there is more direct communication of information from publishers to the national library. Other strands could usefully be incorporated into the model to cover patterns which have been adopted elsewhere in Europe, for example emphasising the role of other libraries and library union catalogues in the information flow.

4.1 Organisations Involved in the Metadata Information Flow

The organisations currently involved in the transmission of metadata have been categorised as :

4.1.1 Publishers

UK examples of publishers substantially involved in a mix of print and non-print publishing and/or with relatively well-developed internal information systems: Oxford University Press, Institute of Physics, Dorling Kindersley, Reed Group (both academic and general trade publishing).

4.1.2 Information services

For example compilers and suppliers of metadata independent of the supply of actual publications.

- **National library bibliographic agencies.**

UK example: the British Library National Bibliographic Service.

- **Trade bibliographic agencies.**

UK examples: Whitaker, Book Data, Bibliographic Data Services.

- **Abstracting and indexing services, typically subject-based.**

UK example: INSPEC (The Institution of Electrical Engineers).

- **Serial contents databases, not subject-based.**

These may be associated with a journal subscription agent or national library document supply service, but have been treated as a separate category for the purposes of this analysis. It includes new 'table of contents' services, often provided as part of larger document supply service. CARL UnCover also fits in here. These services will become more common as a variety of players, including publishers themselves, enter document supply. UK examples: Blackwells, British Library Document Supply Centre.

4.1.3 Suppliers

The suppliers of actual publications who also supply metadata.

- **Booksellers and library suppliers.**
UK examples of booksellers particularly active as (a) users of externally-supplied metadata in their own business: W H Smith, Dillons, Heffers, or (b) onward suppliers of metadata to library customers: Askews, Dawsons, Holt Jackson, JMLS/Blackwells etc.
- **Journal subscription agents.**
The major agents are international: Blackwell Group, Dawsons/Faxon, Swets, EBSCO, etc.

4.1.4 *Libraries*

- **Library union catalogue organisations**
These are organisations which hold union catalogues of their members, and are involved in the supply of records to libraries. They may be commercial, or non-profit companies, or part of the library sector. They are sometimes linked to library management system suppliers (e.g. in the UK such organisations as LASER, BLCMP, SLS, CURL; in Sweden BTJ and LIBRIS; in the Netherlands PICA; and throughout Europe OCLC).
- **Contributing Libraries**
Other libraries contributing to the work of the national bibliographic agency e.g. in the UK other legal deposit libraries

With the increase in electronic delivery of services there has been a blurring of the roles of the various organisations identified above in the metadata information flow, as can be seen by the complexity of the information flow in the diagram. (Annex 1). It seems likely that with the increase in on-line publication and delivery organisations will continue to take on multiple roles with a parallel growth in the complexity of information flow. Creation and enhancement of metadata will be carried out by a variety of organisations, all of whom will be contributing to the 'bibliographic control' of electronic publications.

There are already new organisations emerging which already have a role in the production of information about electronic publications. For example:

- **Authors**
A number of initiatives are underway which aim to encourage authors, or document producers, to 'embed' metadata in their web and other documents.
- **Internet search services**
These may be global commercial services (such as Yahoo, Excite) or selective subject services (such as the UK eLib services OMNI, SOSIG and others, Swedish engineering service EELS, or the Dutch national service NBS). At present these are the chief source of bibliographic control of resources on the Internet. An important feature of the latter services is that they involve informed manual selection of description of included resources.
- **Electronic archive services**
Archive services are collecting, preserving and providing access to electronic material. As part of this venture they create metadata, often of a highly detailed nature, to describe these documents. Examples of such organisations are the Oxford Text Archive, the Essex Data Archive, the Electronic Text Centre at the University of Virginia.
- **Repository collections**
A number of repositories of electronic publications are becoming established, particularly in the academic world. Repositories provide access to current selected material but may or may not have an archival function, e.g. Los Alamos National Physics Pre-print Archive. As the Internet becomes a more mature information environment, many well-managed information repositories will appear, and the supply of metadata will be seen as an important component of such management.

4.2 Metadata Formats Associated with Particular Types of Organisation

The various types of organisation involved in the model use particular metadata formats, and many are

involved in the development of proposed new formats. The formats in the following sections have been identified as being particularly associated with following groups.

4.2.1 *Publishers, Booksellers and Suppliers*

Publishers' bibliographic databases own formats e.g. BIC Manual on Publishers' Bibliographic Databases, draft, 1994. (unpublished, but available on request from Book Industry Communication.) Cataloguing in Publication forms.

EDI standards :

TRADACOMS: Book Trade Price & Availability Updates File, TRADACOMS File Format 108, Version 1, London, Book Industry Communication/Article Number Association, July 1993.

BISAC: BISAC X12 832 Price/Sales Catalog, New York, Book Industry Study Group, 1996

EDIFACT: The EC EDILIBE project, in association with EDItEUR, have defined an EDIFACT QUOTES message to be used for new title notices from booksellers to libraries.

SGML standards: DTDs for serial headers including MAJOUR and SSSH.

Under development: SGML formats for book and serial product information; EDIFACT formats for book and serial product information.

4.2.2 *Libraries, National Bibliographic Agencies, Trade Bibliographic Agencies*

Various MARC formats e.g. UKMARC, IBERMARC, UNIMARC. The MARC records created at various stages of the information flow may be of varying levels of richness. At the Cataloguing in Publication stage there will be a basic record containing few fields (sometimes referred to as a pseudo-MARC record).

4.2.3 *Internet search services*

Proprietary formats: such as those used by OCLC's NetFirst service, internal formats in Yahoo and AltaVista.

Dublin Core

SOIF.

IAFA/whois++ templates.

Resource description messages (RDM).

4.2.4 *Electronic Text Archives*

TEI headers.

EAD.

CIMI.

ISAD(G).

4.2.5 *Repositories*

Under development:

RFC 1807 (A format for bibliographic records) as used in the NCSTRL project.

5. TYPOLOGY OF METADATA

There is a wide variety of metadata formats in existence, and awareness of these formats is becoming more widespread across communities. This offers the opportunity to create descriptions in different and more appropriate formats while retaining the possibility of exchanging data across formats. Awareness of the strengths and weaknesses of the various formats allows choice of the most appropriate format for a particular task. The use of different formats admittedly creates additional effort for achievement of interoperability, on the other hand different formats can be justified if they are needed for particular tasks.

It is clear that the various communities involved in creating and using different metadata formats are strongly attached to their own formats. This is understandable if one keeps in mind such factors as the effort involved in reaching consensus on formats, the skill levels required to apply formats in a consistent way, and not least the heavy investment in existing systems. For these reasons alone it is unlikely that any one format will become dominant. It would also seem undesirable as the existence of a variety of formats allows for choice of the optimum format for use in a particular context.

The provision and exchange of trade and bibliographic data relating to publications has been part of the book world for a considerable time, and has increased with the availability of data in electronic form. Records are created at various stages and places in the process of supplying documents to the reader, they are created to serve different requirements but have overlapping functions. The Newbury seminar held in 1987 (Bibliographic records in the book world: needs and capabilities) considered whether there could be improvement in the 'information flow' whereby the most appropriate record content and format could be sustained throughout the process to meet the various users' requirements (publishers, suppliers, libraries).

One important strand to emerge was the idea of an evolving bibliographic record where through more organised articulation of current record supply, or less likely, through the development of an all through single record system, current requirements might be more efficiently met.

(Bibliographic records: use of data elements in the book world. Lorcan Dempsey, British National Bibliography Research Fund Report, 1989)

In the book world there is still only the beginnings of a more organised evolution of the bibliographic record. The issue of 'a single record system' remains outstanding; at present there is limited integration of records used in the publishing/supplier world and the national bibliography. For electronic publications we need to consider whether it is an ambition to more fully integrate record supply, how far various needs and requirements can be met by the single record system, or whether we accept a more evolutionary system with different record formats available to various users to meet their different requirements.

5.1 Continuum of Complexity and Richness

In order to analyse the various metadata formats it is possible to make approximate groupings based on shared characteristics. Of central importance is the underlying complexity of the format and this suggests a typology of metadata along the continuum from simplicity to complexity. This typology was used in the DESIRE report as an initial means of identifying the characteristics of various formats.

A variety of formats have been placed in this table, positioned along a continuum from simple records (Band One) to complex, rich records (Band Four). The variety of record types identified in the bibliographic control process can be placed on this continuum as shown below.

Band One	Band Two	Band Three	Band Four
Proprietary simple records:	Dublin Core	MARC	ICPSR
NetFirst	IAFA	TEI independent headers	FGDC
AltaVista	RFC 1807		CIMI
Infoseek	SOIF		EAD
Publishers' CIP forms	CIP MARC	EDI messages	
		SGML article headers	

It is possible to extend this model to associate other factors with the position of the format on the continuum. The simplest record formats are used to create relatively unstructured indexes for locating items whereas the most complex records can be used as the basis of sophisticated analysis and navigational tools. Records can be associated with more or less 'rich' retrieval and analysis processes (Z39.50, emerging query routing, text analysis). The bands of records typically have common characteristics in other aspects, for example:

- creation method
how the record is created whether by hand or automatically. This in turn will affect level and type of resource required, and the overall cost of record creation.
- function of record
whether the record is used for location, selection, evaluation, analysis
- complexity of designation
simpler records do not permit complex designation of sub-fields, qualifiers etc
- associated protocols
more complex formats are associated with the more heavy-weight search and retrieve protocol (Z39.50) whereas the simpler formats tend to be associated with directory service protocols.

This pattern of association is summarised in the following table:

Simple	Rich		
Location	Selection	Evaluation	Analysis
Robot generated	Robot plus manual input	Manually input	High level of manual input
Unstructured	Attribute value pairs	Subfields, qualifiers	Highly structured mark up
http with CGI form interface	directory service protocols (whois++) with query routing (Common Indexing Protocol)	Z39.50	Z39.50 (in future with collection navigation)

Simple	Rich
proprietary	generic standards used in information world

Within the context of BIBLINK we need to consider which Band of record is appropriate for further consideration. There may well be a different answer depending whether the requirement is for a CIP type record or a more detailed record.

5.2 Further Scoping Required

In order to make this decision there are a variety of issues that need to be addressed. These include:

- what functions will the record fulfil?
- can we assume that we are not looking for proprietary solutions, but rather for standards supported by international (or at least European) consensus? If so then we can rule out Band 1 formats.
- is the aim to choose metadata format(s) which can also be used for print material?
- what will be the creation method for the records? How much manual effort will be available? Are there cost criteria?
- which Internet protocols will be used? Are there constraints imposed by the technical framework? (This will be handled in WP5)
- what level of designation is required? Will mapping to other formats be necessary and if so how much data loss during conversion is acceptable?

Recommendation

Project partners need to consider these issues and refine the scoping of the project to allow criteria to be drawn up for decision making.

While accepting the need for consensus from partners on these issues, we will assume that previous practice and discussion of these issues will inform future decisions. So for example we assume that there will be constraints on cost and staff available from national libraries and publishers; that we are attempting to identify standard formats that are controlled by authoritative agencies; and that the level of service provided by the national bibliographic agencies will be comparable to that provided for print material (whether this is at a CIP level or at a level consistent with the full record in the national bibliography).

Given these assumptions, the formats from Band One would be rejected as proprietary solutions. The formats in Band Four would also be rejected as too detailed for the service level required, too specialised in nature for general use, and too costly for system maintenance particularly in terms of specialist staff required with skill levels to manipulate and interpret such records.

Predicated on these assumptions as to scope, it is recommended that BIBLINK should concentrate on formats in Bands Two and Three for the exchange format. This does not preclude the possibility that conversion will be required from formats outside these Bands e.g. from more complex formats in Band Four into a simpler format, but that the formats for data exchange would be located in Bands Two or Three.

5.3 Topology of Metadata

An essential aspect of the level of richness of a format is the extent of the content, both in terms of range and depth. The attempt to describe more or less aspects of an object will be reflected in the overall level of complexity e.g. designation, format rules for content. In order to identify the extent of content the elements describing an object can be clustered into groups.

One example of this is given by Bearman who proposes a reference model for business acceptable communication. (Bearman, David and Sochats, Ken. Metadata Requirements for evidence. Available at <URL: <http://www.lis.pitt.edu/~nhprc/model.htm>>). This defines clusters of data elements which

would be required to fulfil the range of functions of a record. The functions of records are identified as the provision access and use rights management, networked information discovery and retrieval, registration of intellectual property, authenticity. The clusters of data elements are defined in six layers:

- Handle Layer
 - registration metadata or properties
 - record identifier
 - information discovery and retrieval
- Terms and Conditions layer
 - rights status metadata
 - access metadata
 - use metadata
 - retention metadata

- Structural Layer
 - file identification
 - file encoding metadata
 - file rendering metadata
 - record rendering metadata
 - content structure metadata
 - source metadata
- Contextual Layer
 - transaction content
 - responsibility
 - business function
- Content Layer
 - content description
- Use history layer.

Bearman's model looks at the record in a wider context than the bibliographic context alone, and it is particularly relevant to our project as it takes account of the business context in which metadata is used. Bearman includes metadata elements which would be appropriate for metadata in the context of publishing and supply, where for example the 'business function' might include sales promotion, price update etc. In a similar way the 'Use history level' would be appropriate for recording changes in electronic documents and assisting authentication.

Within the UK, the BIC/BNBRF Book Product Information project has had as its main aim the identification of the content required for an EDI message to communicate product information through the book sector supply chain for non-serial items. This work has now developed into compiling exhaustive sets of data elements which might be used in this context.

- Line item identity and attributes
 - identification
 - type
 - tradability
 - relationships to other line item
 - publisher

- Work identity and attributes
 - Title data
 - Personal authors, data providers
 - corporate authors
 - description
 - subject
 - relationships to works
- Media specific attributes of a work
 - cartographic scale
 - video duration
- Attributes of a piece
 - media codes
 - dimensions
 - number of pages
 - supply medium
- Attributes relating to terms of trade
 - market area
 - publication date
 - terms of sale
 - price
- Attributes relating to sales promotion
 - Promotional text
 - Promotional images
 - Author promotion

- Message function
- message type (replacement, update etc.)
- effective date etc.

This outline suggests the general categories required to cover the trade and promotional data associated with publications. It also introduces consideration of the level at which objects are described, which may be different in the book worlds and the library world. This is particularly the case in multimedia packs where CDs and books, or more traditionally tapes and books are sold as single line items but may be described as separate items by libraries. Within the BIC study the levels of object are defined as :

Line item: the entity described in the Book Product message, in effect a tradable item

Work: body of literary or intellectual content

Piece: single indivisible physical published item.

Turning to the library world, IFLA has recently issued a draft for comment of its study of the functional requirements of bibliographic records. This study attempts to identify and define objects of interest (or entities) for users of bibliographic records i.e. what information the user expects to find in the record, and how that is used. As regards the intellectual content described by a particular bibliographic record, can be viewed at different levels, each of which can be related to different information seeking behaviours from the user. These entities are:

Work: distinct intellectual or artistic creation

Expression: realisation of the work in text, sound, music, image etc

Manifestation: physical embodiment of an expression in book, sound recording etc

Item: single example of a manifestation

Within the context of BIBLINK it would seem there needs to be a shared view among publishers and national libraries as to what 'entities' or 'objects' are of interest. Traditionally libraries have dealt with printed material to a large extent at the manifestation level e.g. particular editions of a book. As electronic formats offer increasingly varied manifestations, then creation and re-use of metadata relating to the 'work' will take on more emphasis. At a different 'entity' level, it may become desirable to describe collections rather than individual items e.g. for web-sites containing varied and changing information. The problem of the identification of different manifestations and expressions is perhaps more familiar to publishers who are accustomed to dealing with compilations, re-prints, new editions and other 'bundling' of works. The on-going work by BIC to formulate a non-serial DTD has recognised this in the proposed layering of data elements.

Recommendation

As part of the consensus building process, publishers and national libraries identify the objects and relationships which need to be represented in metadata describing electronic resources.

6. METADATA REQUIREMENTS OF NATIONAL LIBRARIES

As part of this work package a questionnaire was circulated to five national libraries: Bibliotheque Nationale de France (BNF), Koninklijke Bibliotheek of the Netherlands (KB), the Norwegian National Library (NB), Biblioteca Nacional (BN) in Spain and the British Library (BL). The full responses are attached in Annex 2, and a consolidated report is included in this section. The questionnaire was designed to gather some initial information about the cataloguing practices of the various national libraries regarding electronic publications, and their future requirements.

The responses appear to support the view that the metadata format(s) adopted for electronic resources are consistent with those adopted for conventional print material. There are obvious benefits to existing national bibliographic agencies if the full range of document types can be handled in a system with a common structure. However if separate services were to be provided solely applicable to electronic resources this would no longer be the case.

The responses indicate that libraries wish to create full records describing objects at a detailed level in terms of bibliographic content, terms and conditions, access, subject content, location etc. They intend to create records in the various flavours of MARC currently used for print publications. They intend to apply detailed cataloguing rules to the content of the records. The records created in this way are envisaged as forming the basis of a range of services involving selection, retrieval and access, record supply and preservation.

As part of the consensus building process we need to establish libraries' requirements as regards the level of detail supplied by publishers. Are libraries seeking detailed records or is the intention to add value by enhancing simple records? Do libraries want detailed records embedded in electronic publications for use at the cataloguing stage or would they prefer simple records to be supplied in advance of publication? Possibly libraries will only ever want simple records for those items which are not legally deposited or for particular categories of publication.

The timing within the publication process of supply of metadata from publisher to library has a crucial impact on the format of metadata. For off-line publications from 'traditional' electronic publishers, national libraries may be wanting to use information received from the publisher in advance of gaining access to the object. In this case libraries are looking for the equivalent of a CIP system. Either there could be a one-off supply of 'minimal set' metadata supplied pre-publication, or additional detailed metadata could also be supplied as it became available. For many newer web based publications speedy creation of metadata is of the essence, the aim is to minimise the delay between appearance of the electronic resource and provision of metadata.

In this context Lenart identified significant factors in the supply and demand of records as:

- delay in supply
- quality of records
- subject indexing
- legal limitations imposed on re-use of records
- charges payable to rights holders.

(Models for the provision of national bibliographic services in Europe : final report / by Michele Lenart ; translated and edited by the British Library. - Luxembourg : European Commission, 1996).

These factors are worth considering in the context of electronic publications and can be interpreted as:

- at what stage of the publishing process are records required by national libraries?
- at these identified stages what quality of record is required? (The required level of detail and granularity may differ)
- what added value is required at the various stages of the process? (For example inclusion of subject indexing, adherence to cataloguing rules, adherence to authority files may vary)
- what are the implications for ownership of records? (If publishers supply detailed records they may wish to retain ownership of these records).

As regards this last point, the business model will differ significantly from country to country. Lenart identified re-use and copyright issues as significant in the following countries: Belgium, Greece, Ireland, UK, Portugal and Italy. However the cost implications of other factors (level of detail, timing of record supply, added value) would be significant in all business models.

Another issue for consideration is the disparity between publishers and libraries in their use of rules for formulation of content. Typically publishers do not follow cataloguing rules for content even in detailed SGML headers. Will complex information supplied by publishers be of benefit to libraries if it needs to be re-formulated to adhere to cataloguing rules? As part of the refinement of scoping project partners need to consider how far they will guide publishers as regards rules for formulation of the content of the information they supply.

It is worth noting that three libraries (BNF, KB and NB) indicate they are using the Dublin Core elements as a way of exploring means to catalogue electronic resources.

The advantages of embedded metadata as opposed to 'free-standing' metadata are alluded to, one significant aspect of this is authentication. We will leave this issue to the authentication work package.

6.1 Responses to the Questionnaire

All of the five libraries have had some experience in cataloguing off-line publications. These are deposited (either legally or voluntarily) in the Netherlands, France, Spain and Norway. In the United Kingdom consideration is being given to extending deposit laws to non-print items. KB and NB have started to create records for on-line publications.

6.1.1 Cataloguing format and cataloguing rules used currently

Q1. Please indicate the cataloguing system you use and the cataloguing rules you apply.

Responses indicate that each country uses its own cataloguing rules, all based on ISBD, and its own flavour of MARC, which will vary more significantly.

KB	PICA	PICA rules are ISBD based. Guidelines for electronic documents are under development.
BL	UKMARC	AACR2.
BN	IBERMARC	Spanish Cataloging Rules (based on ISBD).
BNF	INTERMARC	AFNOR (National French Standardisation body) standards for cataloguing rules. Plus ISBD(CF) based rules for electronic documents.
NB	BIBSYS-MARC based on USMARC	Norwegian Cataloging Rules (based on AACR2) plus a Norwegian supplement for off-line machine readable files and an OCLC guide for on-line.

Q2. What data elements do you record for an electronic publication *in addition* to those you record for a traditional publication?

Q3. Which MARC (or equivalent) fields do you use to hold the data element

Q4. What data would you like to include in the record that you cannot find, or have difficulty finding from the publication?

Q7. What metadata elements do you consider will be required for electronic publications?

Examples: description of content (including mention of sub-units, contents pages); relation to other documents

This table lists all the data elements mentioned in response to questions 3, 4 and 7 as well as question 2. It therefore includes:

- elements *already* recorded in cataloguing electronic publications
- elements that *should be* recorded but can be difficult to find
- elements it is felt *will need to be* recorded in the future.

Where an element was mentioned by only one library this has been noted in the last column.

Data Element	Comment	
personal author	Person primarily responsible for the intellectual content.	
other contributors	Statements of responsibility for multiple contributions.	
corporate author		
definitive title	Variants of the title can appear on boxes, accompanying information and internal sources - these often conflict.	
unique identifier	e.g. ISBN.	
place of publication		
publisher	Agency responsible for producing the publication.	
host	Agency making the publication available - e.g. SURFnet.	
date, year	Date of publication.	
exact date of issue	To distinguish updates in electronic resources.	
price	For off-line publications.	
language		
edition	Particularly significant for on-line publications.	
update information	"	
version information	"	
general material designation	e.g. [computer file].	
specific material designation	e.g. tape, diskette etc.	
type of computer file	e.g. data, program etc.	
file characteristics	e.g. size, number of records contained.	
additional information	e.g. sound, image, text, multimedia etc.	
relationship to printed versions	Existence of printed versions e.g. a digitised novel.	
subject keywords	Need for controlled vocabulary. Cataloguers may use publishers suggestions as a guide.	
legal deposit number		
classification		BN
description of content	More detail than keywords alone would assist selection and acquisition by users of BNB. Summary, tables of contents.	BL (BNB)

<i>Additional data for on-line publications</i>		
availability	Free of charge or by account.	
terms and conditions		
login name and password		
file format	HTML, pds, ps, wp51, etc.	
URL		
file name		elements defined by KB
path		"
number of files, bytes		"
compression format		"
type of connection		"

Data Element	Comment	
port number/protocol		"
gopher type		"
name of computer - host		"
IP address computer - host		"
location - host		"
mail address - host		"
mail address - person		"
URL and other details as above for images linked with publication	To present an image which is linked with the document as part of the bibliographic description.	"
service provider		
links	A listing of links to other documents.	NB
unique title		NB
duration of availability at given URL		BNF

<i>Additional data for serials</i>		
frequency	How often it will appear.	
regularity	Information relating to the production of new articles. Will they only appear at with the next issue, or as and when they are ready for publication?	BL
data relating to articles	Information relating articles to the journals they appear in.	
abstracts		
contents page data	See also, 'description of content' in the main table.	BL
<i>System requirements</i>		
Statement indicating if specified system requirements are preferred or required.		
Agreement on standard place in publication and/or in accompanying documentation for system requirements.		
off-line	Processor, memory, operating system, application software, monitor, cards, peripherals etc.	
installation and de-installation information		
on-line	Internet browser, viewer, telnet client, FTP client, WWW client etc.	
<i>Other related data</i>		
Amount of time taken to install	This information will be important for access to the documents by the end-user - it needs to be taken into consideration for the access service.	BNF
local address	For off-line or remote material made available on a local network.	

Q3. See 2 above

Q4. What data would you like to include in the record that you cannot find, or have difficulty finding from the publication?

The individual elements mentioned in response to this question have been included in the tables above. The following notes include some additional remarks.

BNF, BL,
KB Difficulties finding the source of the description. In electronic publications the information is scattered throughout the document and can take some time to find.

BL, BN,
BNF Statements of responsibility: a CD-ROM can have many different functions : script, developer, infographist, designer, title manager, music, etc.

KB, BN, BL,
BNF, NB Technical data and system requirements: for installation and de-installation for cataloguing; for access now; for access in the future. Distinguish between *required* and *preferred* hardware and software.
It would be easier if all this was presented in a standard format and location in a publication.

KB, BNF,
NB Terms and conditions/access information.

- BL Publisher name and location.
- BL Correct title - variants on boxes, accompanying information and internal sources often conflict.
- BL, BN Version/edition information. (Is a serial a first issue?)
- BNF Precise date of publication.
- NB The 'size' of an on-line document. What (and how many) files/records does it consist of? A listing of links to other documents.
- BN ISBN and legal deposit numbers are not shown on disc labels or elsewhere.

Q5. How have you resolved or attempted to resolve the difficulty?

All the libraries had some experience of cataloguing off-line publications although at some this was still at the developmental stage, and for some publications, relied on external sources. Off-line publications are treated very much as printed books or serials with extra data added relating to system requirements.

- K establishing test beds.
- B
- BL developing rules to apply to problems as cataloguing rules do not let you assume anything.
- B contacting publishers for further information.
- N
- N contacting other libraries to see how they have resolved the problem.
- B asking advice of IT personnel.

Q6. What library services will use metadata for electronic publications?

- KB OPAC (local retrieval system).
 - Union Catalogue (NCC) (national retrieval system + ILL).
 - Delivery of electronic resources e.g. published by universities (WebDOC).
 - List of electronic publications as a by-product of the national bibliography.
 - (possibly in the future) alerting services.
- BL Facilitating access and record supply
 - Selection
 - Acquisition
 - Preservation and archiving
- BN Most library services will use metadata.
 - A problem to be considered with deposited electronic publications is copyright and a possible unfair use.
- BNF National Bibliographic Agency both for legal deposit and National Bibliography.
 - Specific bibliographic by-products for electronic documents
 - Services in charge of access to the collections.
- NB All library catalogues
 - The metadata will probably have to be refined to meet the requirements of the national bibliography.

Q7. What elements will be needed in future? See 2 above

Q8. Will the records for electronic publications need to be integrated in your existing systems?

- KB Yes, the records of electronic publications produced in the PICA format are in the same database with the other title records. It must be possible to have both kinds of records in the OPAC's of the library and e.g. in files of the National Bibliography. A total integration is required.
- BL As far as practicable and to the same basic bibliographic and format standards. The British Library is aiming for uniform access to its collections.
- BN Yes. The National Library wants to have a single database.
- BNF Off-line electronic publications are already integrated in our system. For on-line, it will be the same. We will have a unique multimedia integrated system for the OPAC and national bibliography. This requirement is very important because we will adapt the existing system to new electronic documents but we will not completely change our system or our format in the next five years.
- NB Yes.

Q9. Will the metadata need to be manipulated by particular protocols e.g. Z39.50, EDI, etc.?

- KB PICA exploits since the end of 1995 a Z39.50 entry to the central database GGC. In view of this it is important to investigate the pros and cons of manipulating metadata by the Z39.50 protocol in comparison with the use of the PICA protocol.
Nowadays we buy tapes with bibliographic records which are integrated in the common database maintained by PICA. As output, tapes are sent to our CD-ROM production company in the United States. Possibilities to use FTP for transfer of the records of the national bibliography for the CD-ROM are investigated.
- BL Probably MARC, Z39.50, HTML.
- BN Yes. At least, Z39.50, ISO 10162/10163, EDI, and any other protocols established by the DG.XIII.
- BNF Yes, the catalogue will support the Z39.50 protocol. We plan to give access possibilities to other libraries to the BNF national bibliography records via FTP.
- NB Yes, certainly by Z39.50. And by other relevant protocols (even if we do not use them e.g. the EDI protocol).

Q10. Further comments about experiences in cataloging electronic publications.

- KB Off-line electronic publications are catalogued and published in the national bibliography. The cataloguing of CD-ROM and diskettes does not present specific problems. Except for the difficulty in some cases to judge whether a hybrid publication (book + diskette) is mainly a printed or mainly an electronic publication.
- Concerning on-line resources, especially home pages: At this moment three cataloguers are making bibliographic descriptions for home pages. One for the scientific collection, one for the union catalogue and one for the national bibliography. Further, since a common cataloguing system is used they can also extract records from the common database and they can see how cataloguers in other organisations, especially one university library catalogue e.g. home pages. The problems with home pages are: author unknown, publisher unknown, date of publication unknown, change or even disappearance of the URL.
- BL In terms of physical carriers, a cataloguer may be presented with any number of formats - CD-ROMs, floppy disks of two different sizes, magnetic tapes, cartridges, all of which may be specific to different platforms (PC, Mac, Acorn, etc.). It will be necessary to be in possession of the equipment and software needed to access the publication.
It takes considerably more time to create a record for an electronic publication due to: 1) installation and de-installation 2) procedures to give internal access may entail reading

more of the document than in paper publications.

One of the systems issues raised by the project is the difficulty in having simultaneous access to the WLN cataloguing system, and the item in hand. The ability to "hot-key" between the item's title screen and the cataloguing form screen is essential.

BN We have little experience of this type of publication but we are very interested in all problems with this material.

BNF BNF has a solid experience for cataloguing off-line electronic publications. The on-line topic is very important to us and we will benefit from BIBLINK results before starting a national project.

The French national bibliography is based on the legislation for legal deposit which includes off-line electronic publications. The problem is to include progressively on-line publications but then we come back to the limits. What do we include: electronic journals, home pages (on which criteria etc).

The evolution towards cataloguing "articles" is also an important issue as we do not include articles in the national bibliography.

Q11. Wish list - anything else?

KB For the National Bibliography and other metadata products and services of the library, the consistent and user-friendly presentation of records on screen or on paper is very important. In the international library community a proposal for an ISBD(CF), the presentation of metadata of computer files, has been commented upon. In December 1996 the final version is expected. In the meantime we have defined ourselves an 'ISBD'-like presentation for 'On-line Resources' and PICA has implemented this in its system.

Linking between different versions of electronic publications.

Links between journal and articles of the journal (whole/part relationship) should also be realised in the system.

BL Every item (off-line) should have a de-installation file.

There should be a standard place and format for including installation instructions in the user handbooks accompanying the item.

NB The quality of the metadata elements should be ensured, especially by controlled vocabulary, e.g. name of persons and institutions, subject keywords.

7. IDENTIFICATION OF METADATA FORMATS FOR FURTHER CONSIDERATION

A detailed review of metadata formats was undertaken as part of the DESIRE project and is now available as a deliverable (draft). We do not intend to repeat the detailed examination of the full range of formats contained in that report. Here we shall attempt to summarise characteristics of those formats which might be used as solutions within the context of BIBLINK. We will also include consideration of publishers' formats which were not included in the DESIRE overview.

The various categories of players involved in the information flow are associated with different metadata formats as described in section 6 above. From this list we have chosen particular formats as suitable for the requirements of both publishers and national bibliographic agencies, and as likely options for use within BIBLINK.

The criteria for selection are:

- familiarity of use within the relevant sectors
(traditional publishers, web publishers, national libraries, booksellers and other bibliographic agencies)
- appropriate level of detail
(note that according to the present scoping of the project it is unclear what level of detail is required. This will be dependent on such factors as whether CIP records are required, and whether the resource being described originates from traditional publishing or web publishing. For this reason formats at different levels of detail are included)
- ability of the format to accommodate description of electronic resources
(we do not intend to consider formats where there is no signal of adoption of elements appropriate for electronic resource description)
- the format has some grounding as an international standard
(There is an intention among the stakeholders to achieve acceptance as an international standard in whatever forum is appropriate)
- the cost of creating the record is appropriate to the environment
(It is assumed that detailed descriptive records, as categorised in Band 4 section 5.1, will not be created by publishers or national libraries. However we will assume that the information in such records may well feed into the process. So publishers such as archive services who create records in CIMI, IPCSD or EAD format may wish to map these records into another preferred format to feed them into the process of bibliographic control).

The formats chosen for further investigation are:

Simple Internet descriptions:

- IAFA
- SOIF/RDM
- Dublin Core.

Publishers formats:

- MAJOUR: Modular Application for Journals, European Workgroup for SGML, 1991
- SSSH: Simplified SGML for Serials Headers, London, Book Industry Communication and PIRA International, 1996
- EDIFACT formats for book and serial product information.

Archive formats:

- TEI (and its relation to CIMI, EAD).

8. DESCRIPTION OF SERVICE AND TECHNICAL CONTEXT FOR METADATA FORMATS

The following headings will be used to analyse the selected formats:

- community of use
- what is the level of maturity of the format?
- Is there a control agency?
- how widespread is deployment?
- overview of technical issues (e.g. associated protocols)
- future path for the format (development plans)
- content
- extensibility (will the format allow for inclusion of new data elements and description of new object types?)
- rules for formulation of content

8.1 Simple Internet Descriptions

8.1.1 Dublin Core

8.1.1.1 Community of use

The strength of the Dublin Core metadata format is that it is built upon the basis of international consensus. A range of interested parties from different professional backgrounds and subject disciplines have contributed to the development of the format. There has been high commitment and involvement from a range of professions (publishers, computer specialists, librarians and information workers) and sectors (library utilities, software producers, service providers, libraries). The motivation progressing Dublin Core has been to reach a consensus among stakeholders on a minimal resource description which can be used for the benefit of all involved in the creation, search and retrieval of electronic resources.

In the context of BIBLINK, Dublin Core is significant as 'web publishers' have been involved in the consensus building process. In addition library and information personnel from Europe and the US have been involved in the workshops and discussion lists. The Library of Congress are supportive of the consensus building work and have investigated use of the format.

The Dublin Core is positioned as a simple information resource description. Importantly it also aims to provide a basis for semantic interoperability between other, probably more complicated, formats. A third target use is to provide the basis for a format for embedded metadata i.e. metadata contained within the body of the resource. It is in fact in this third role that most progress has been made, with a high level of interest from those involved in automatically harvesting metadata from HTML documents.

8.1.1.2 What is the level of maturity of the format?

A series of workshops mark the significant steps in the history of development of the format:

- OCLC/NCSA Metadata Workshop (March, 1995)
 - original definition of Dublin Core element set
- OCLC/UKOLN Warwick workshop (May, 1996)
 - definition of syntax for Dublin Core
- W3C Distributed Indexing and Searching Workshop (May, 1996)
 - refinement of HTML syntax for Dublin Core
- CNI/OCLC Image Metadata Workshop (September, 1996)
 - refinement of DC element set to accommodate requirements of image description
- proposed DC meeting in Australia Spring 1997
 - possibly this will mark agreement on version 1 of the format.

The issue of a draft User Guide in September 1996 and input from the ERCIM meeting (October 1996) have also stimulated much discussion and some re-working of the original definitions of data elements. As yet the Dublin Core format is not stable, but the level of international involvement means that once agreement is reached on the format there are likely to be several implementations. The strong backing of the format by OCLC would imply that they see Dublin Core as the format of choice for provision of 'Internet publication' services i.e. as a format for their NetFirst electronic publication search service, and as a format for record supply and exchange.

8.1.1.3 Controlling Agency

As yet no control agency has been established, but the level of international involvement should assist in agreement on establishing an authoritative agency.

8.1.1.4 How widespread is deployment?

A number of projects and initiatives are in early stages.

In Europe the Nordic Web Index is committed to using Dublin Core. Within the UK there is considerable interest in pilot implementations as part of ROADS and the AHDS service.

There is a joint project between the National Libraries of Australia and New Zealand, the National Document and Information Service (NDIS). In this project Dublin Core is used as a means of achieving semantic interoperability by mapping core elements from disparate complex records onto the core element set. Within this project the Dublin Core elements have been used as the core search attributes for all records, in effect the intersection between the various databases included in the service. This has allowed flexibility in the use of semantics across databases, with mapping of other 'search fields' to the Dublin Core set.

DSTC in Australia is using the Dublin Core in the Research Data Network Co-operative Research Centre project for resource discovery. In this instance Dublin Core is used as a simple record format for the storage of metadata, and for searching and retrieval purposes.

Within other contexts such as the Archaeology Data Service (ADS) in the UK, the embedding of Dublin Core in documents is seen as an important means to create metadata, and to enable that metadata to be automatically harvested. The Arts and Humanities Data Service, of which ADS is one of the groups, is investigating the use of Dublin Core to provide a general catalogue to its various sites, with points directly at resources themselves or at richer metadata formats (e.g. TEI).

8.1.1.5 Overview of technical issues

At the present time there are moves to create Z39.50 attributes corresponding to the Dublin Core set. However at the same time there is a wider discussion taking place within the Z39.50 implementors Group as to how Z39.50 should take attribute sets forward. There is a suggestion that the control of attribute sets should move out of the standard itself and be controlled by agreement within other domains. Cross currents across both areas of discussion may mean agreement will take a little time. Dublin Core records could be adapted for use with the directory service whois++.

There are proposed SGML and HTML syntaxes for Dublin Core. However there are suggestions that other syntaxes could be used e.g. PICS.

8.1.1.6 Future path

Dublin Core is designed as a means for 'publishers' and authors to provide metadata at the point of mounting information on the Web. It is in the interest of these publishers to make metadata available which can be harvested by commercial and selective search services as a means to ensure their publications become publicised. Similarly search services are likely to promote a standard format for embedded metadata to make the harvesting process more accurate. Whether Dublin Core will be the format favoured, or whether the web browser suppliers favour their own format remains to be seen. The Dublin Core is also being used as a simple format for third party creation and as a basis for semantic interoperability between richer formats. It is therefore situated precisely in the CIP area.

8.1.1.7 Content

Dublin Core is positioned as a simple Internet description, and the intention is to provide limited, minimal information about a resource. The core set agreed at the first Dublin Core workshop in 1995 consists of:

- Title
- Author
- Publisher
- OtherAgent : person(s) such as editors or transcribers who have made a significant intellectual contribution to the work
- Date
- ObjectType: genre of the object e.g. home page, poem, working paper
- Language
- Subject
- Coverage: spatial location and temporal duration
- Form
- Source: object from which this object is derived
- Relation: relationship to other objects
- Identifier: string or number used to uniquely identify this object.

Since the original core set of elements was defined, there has been considerable debate on whether adjustments should be made to the core set, whether to add additional elements (e.g. a description or abstract) or to combine existing elements (e.g. to combine OtherAgent and Publisher and/or Author). Indications are that this debate will be resolved early in 1997 with the addition of another element: Description.

There is an intention to elaborate on the original simple set of elements to allow for a richer element set, using qualifiers to specify the type, scheme and role of elements. Such qualifiers might either refer to external schemes to be applied for processing e.g. Author (scheme USMARC); or they might specify more precise information about the element e.g. OtherAgent (role editor). The attempt to bridge the gap between simple and sophisticated could potentially cause problems with interoperability unless there is close control of the definition of such qualifiers.

8.1.1.8 Rules for formulation of content

No particular rules are specified but by using the qualifier 'scheme' one can indicate whether the content of a particular element complies to an external scheme e.g.

Author = value (scheme = USMARC)

This does provide the flexibility for different schemes to be applied to different elements.

8.1.1.9 Extensibility

Dublin Core was originally targeted at describing 'document like objects'. Such objects were not closely defined but it could be argued that the emphasis within discussions was on text based resources. More recent discussion in a wider context has highlighted the need for description at the collection level (whether collections of web pages, or collections such as archives). Dublin Core is not designed to allow for navigation between collection level and individual items.

The latest CNI/OCLC Image metadata Workshop, October 1996, has proposed some amendments to

the element set in order to accommodate description of images. This would indicate there may be further amendments required to describe other object types (e.g. sound).

8.1.2 IAFA templates

8.1.2.1 Community of use

IAFA (Internet Anonymous FTP Archive) templates were originally used to provide some bibliographic control over FTP archives. When the templates were designed many organisations were making information available using FTP archives, and the IAFA template allowed the archive to be described from a number of aspects (documents contained there, logical collections of documents, site and configuration details and so on.). The original design came from the IAFA working group of the IETF (Internet Engineering Taskforce). With the emergence and increasing information being made available on web servers, the IAFA templates have been used to fulfil a similar function and they are now used to describe a variety of networked resources.

8.1.2.2 In what service areas are the implementations?

At present the templates are used among the higher education community in the UK as the record format in selective search services. As part of the EC project DESIRE, their use is proposed for subject based gateways covering the Nordic countries and the Netherlands.

8.1.2.3 Controlling Agency

It has been proposed that UKOLN act as a registry for the format for UK users. Within the international community Bunyip are at present fulfilling this role. There is an issue as to how the format will be controlled over time, and who will participate in change control.

8.1.2.4 What is the level of maturity of the format?

The format has been in use for approximately two years. It has proved a successful format for creating simple descriptive records, and is the basis of a number of production systems. The format has been proved by use, and various amendments and changes have been introduced as a result of implementation experience.

8.1.2.5 How widespread is deployment?

There are several implementations which use IAFA/whois++ templates. The first implementation was the ALIWEB service which enabled searching of FTP archives, in effect a forerunner of today's Internet search services. Current implementations involving bibliographic descriptions include a number of production services: SOSIG, ADAM, OMNI, EEVL (variant), NetEc, IPCA.

8.1.2.6 Technical considerations

IAFA templates are associated with the whois++ directory service protocol. There is now a lot of activity in development of directory services based on whois++ protocol, although this is chiefly concerned with 'white page' applications (i.e. names and addresses).

8.1.2.7 Content

The original IAFA templates contain simple bibliographic descriptive elements, administrative metadata, and the means to describe access and location. Within the ROADS implementation additional elements have been added to enable subject headings and subject scheme to be specified.

8.1.2.8 Rules for formulation of content

Some rules for content are specified in the original guidelines, but they are rather patchy. These require elaboration. Within the ROADS project there is a commitment to formulate simple cataloguing guidelines as an aid to those creating simple Internet description records.

8.1.2.9 Future path

Among ROADS users there is a requirement to indicate relationships between web pages. One way this might be done is to create template types for different levels of object i.e. web sites, document collections, individual items. Another suggestion is to use the category field within the template to indicate relationships. The next version of ROADS will explore these possibilities. In addition ROADS will be introducing some internationalisation of the ROADS templates in the next version, to support the use of ROADS in the DESIRE project. The intention is to facilitate multi-lingual descriptions and subject headings.

8.1.3 SOIF / RDM

8.1.3.1 Community of use

The SOIF (Summary Object Interchange Format) is a record format used by the Harvest software. Harvest software was developed at the University of Colorado at Boulder, and is distributed by them as shareware. It is documented at <http://harvest.cs.colorado.edu/Harvest/>. The Harvest architecture includes a Harvest gatherer designed to collect data regarding Internet documents, and a Harvest broker which is designed to enable users to search these records. The Harvest gatherer can generate SOIF records from documents held in a variety of formats (SGML, HTML, PostScript, MIF and RTF). In order to solve problems of limited resources and seemingly limitless web publications, many organisations and services are considering the benefits of generating records using robots. The Harvest software is one example of such a system. Because of its availability it is being considered for use in a number of projects (ROADS, DESIRE).

There is great flexibility in the data elements which can be used, which means it can be useful for a variety of applications, but there is no guarantee of interoperability. Each Harvest broker can support any attributes that are required by the data which it describes, although a set of common data elements has been defined to promote interoperability.

8.1.3.2 In what service areas are the implementations?

The main area of implementation is for Internet search services: most SOIF records are generated by robots, although as they are based on simple attribute:value pairs they can easily be generated by hand. SOIF records can also be used as an aid to creation of other metadata formats.

8.1.3.3 Controlling Agency

There is no identified controlling agency.

8.1.3.4 Content

A broker can support different attributes, depending on the data it holds. Often brokers will hold the full text of documents as well as metadata. A list of common attributes is provided in the documentation as follows:

Bibliographic type attributes:

- Abstract
- Author
- Description
- Title
- Type
- URL
- File Size
- Full-Text
- Keywords
- Last-Modification-Time.

Administrative metadata attributes:

- Gatherer-Host
- Gatherer-Name
- Gatherer-Version
- MD5 (checksum)
- Refresh-Rate
- Time-to-Live
- Update-Time.

8.1.3.5 Rules for formulation of content

Rules for content form are not specified and there is no specified way to indicate whether particular rules or schemes have been applied to content. Any agreement on rules for content would need to be made between co-operating parties.

8.1.3.6 How widespread is deployment?

Harvest has been widely taken up within the academic community and as a basis for search services. Of significant importance has been the recent adoption of Harvest technologies by Netscape. In 1996 Netscape announced they would use SOIF as a basis for their Catalog Server product.

8.1.3.7 Technical issues

In a significant extension to the Harvest architecture, Netscape are working on 'Resource Description Messages' which provide a framework for the creation and communication of metadata. Resource Description Messaging (RDM) is a messaging format which can be used as the basis of a query syntax. It allows for exchange of record descriptions and is particularly designed for use with SOIF records. The client can send a RDM request in order to select resource descriptions. RDM also allows the client to access a schema definition to which resource descriptions conform (data type and format of attributes), in addition RDM supports access to a taxonomy description for the resource descriptions (classification scheme), and a server (catalogue service) description.

The combination of SOIF and RDM means once a repository of SOIF records exists, the server can export it as a whole or on the basis of selection using RDM to retain the structure.

8.1.3.8 Future path

The involvement of Netscape clearly has major potential significance. In addition the widespread use of SOIF within the Internet search services means there is familiarity with the format among an international technical community. It remains to be seen how this work develops.

8.2 Publishers Formats

8.2.1 MAJOUR SGML DTD for journal article headers

8.2.1.1 Community of use

MAJOUR (Modular Application for Journals) was developed by a European Workgroup on SGML linked to the Scientific, Technical and Medical Publishers group, Amsterdam and published in 1991. The group consisted of Elsevier Science Publishers, Kluwer, Springer, Thieme, Fachinformationszentrum Karlsruhe, Stuertz, MID/Information Logistics Group and Satzrechenzentrum Berlin.

8.2.1.2 Maturity/consensus

No full consensus has been reached among journal publishers on an agreed DTD for article headers. Journal publishers apparently have not been able to agree on MAJOUR as anything more than a basis for individual variants.

8.2.1.3 Controlling Agency

No controlling agency has emerged.

8.2.1.4 How widely deployed

Many major serial publishers, principally in the STM field use their own individual adaptations of the MAJOUR DTD. The DTD is viewed more as an exchange format than as a way of storing records internally in publishers' databases.

Note there is an international standard for article headers, ISO12083, based on work by the American Association of Publishers but this was considered inadequate by many publishers and never widely deployed. The American Association of Publishers and the European Physical Society developed this standard method for marking up scientific documents, and it is particularly developed for mark-up of mathematical information.

8.2.1.5 Future path

The MAJOUR DTD tends to be used as an interchange standard rather than a standard for storage. Publishers tend to use DTDs for their internal databases which can be converted to MAJOUR for interchange purposes.

8.2.1.6 Content

Initially covered the "header" to a journal paper, with a statement of intent to cover the text and end matter in further DTDs. The aim was to provide a common electronic language shared by authoring bodies, publishers, typesetters, and publishing and database access software. The DTD for the article is detailed and aims to be comprehensive.

8.2.1.7 Extensibility

Extensible in principle, but depends on there being an effective development and control agency.

8.2.2 SSSH SGML DTD for article headers

8.2.2.1 Community of use

SSSH (Simplified SGML for Serial Headers) is used by journal publishers in all fields, particularly for the purpose of communication from publishers to users. The aim of the SSSH is to harmonise MAJOUR with OASIS' requirement for a simpler set of elements.

8.2.2.2 What is the level of maturity of the format?

The format was developed for BIC by Pira International and was published in 1995/6. It builds on and attempts to unify the MAJOUR approach with different requirements identified by the OASIS (Organisation for Article Standards in Science) group of publishers. OASIS demanded a simpler DTD than MAJOUR and in 1995 published a minimum set of elements consisting of 24 fields. Reaching consensus on a single DTD for article headers is difficult, in that some publishers specifically choose not to supply particular data elements.

A revised edition is in the pipeline. The main purpose of the new version (SSSH2) is to enable the inclusion of alternative article identifiers. The Publisher Item Identifier (PII) is now included. This was

developed by Elsevier Science and adopted by other publishers such as the American Chemical Society, the American Institute of Physics, the American Physical Society and the IEEE. It is anticipated that other alternative identifiers will be added in future revisions.

The publication of the revised version included the addition of the ISO special character entity set for Mathematical Script Characters, and minor changes to the DTD to bring it into conformance with the Reference Concrete Syntax of the SGML standard.

8.2.2.3 Controlling Agency

BIC will be the control agency in association with other relevant bodies

8.2.2.4 How widespread is deployment?

The DTD is being used by some of the UK eLib journals, and some major publishers have indicated they will move to SSSH (Springer, Elsevier, Kluwer)

8.2.2.5 Technical considerations

This DTD offers a fuller character set than MAJOUR including ISO special character entity sets.

Version 2 includes:

- addition of the ISO special character entity set for Mathematical Script Characters
- minor changes to the DTD to bring it into conformance with the Reference Concrete Syntax of the SGML standard.

8.2.2.6 Future path

The aim will be for BIC to build, maintain and develop a comprehensive and consistent set of SGML DTDs for all publication types.

8.2.2.7 Content

Currently consists of article header. It differs from MAJOUR in a number of aspects but where possible attempts compatibility. Compared to MAJOUR it has simplified tags, simplified author affiliations and allows for SICI article identifiers. Parameterisation allows MAJOUR definitions to be restored if desired.

8.2.2.8 Extensibility

Both technically and organisationally capable of extension, through BIC as control agency.

8.2.3 BIC SGML DTD for non-serial publications ("books")

8.2.3.1 Community of use

Intended for use throughout the "book" supply chain where detailed bibliographic, trade and promotional information must be communicated. Although this is commonly referred to as a book DTD it is intended to cover a wide range of non-serial electronic and print publications. The intention is for the DTD to accommodate a variety of media, but that the first version should be targeted at printed material.

The aim is to define a DTD suitable for use by a wide range of publishers large and small. It is intended to define a format which can be used for a variety of functions including internal databases, book promotion, transactions in the trade, and provision of bibliographic information. It is acknowledged that although there will be an attempt at exhaustivity in inclusion of data elements, a minimum list will need to be defined for smaller organisations, or those who do not wish to supply detailed information.

The inclusion of several elements relating to trade information means that the resulting records would be useful for resource selection purposes.

8.2.3.2 What is the level of maturity of the format?

The DTD is under development with the aim of completion in 1997. Work on this DTD grew out of a need recognised for some time in the library and book world for a fuller product description which would include more trade information than the traditional bibliographic record, and also from the need for a record which could match the 'line item' object level required by publishers and book suppliers. This was recognised in David Martin's original report which was the basis of the further work on this DTD. This original work included a survey of 49 BIC member organisations to contribute details of the data elements that they used on their own internal databases.

EDItEUR have agreed to review the DTD when complete and some members of that body will provide input. Some US publishers, particularly of CD-ROMs, support the development (Bowker, Ingram, and Baker & Taylor).

8.2.3.3 Controlling Agency

BIC will be the control agency in association with other relevant bodies.

8.2.3.4 Technical considerations

There is a recognition that the relational database approach will provide advantages for internal databases, but that electronic communication between systems requires a record structure.

The book DTD will certainly have as wide a capability on character sets as SSSH.

8.2.3.5 Future path

Development plans: Initial release to cover at least non-serial publications on paper, in a structure designed to cover all media, including electronic and hybrids. Options for implementation include transmission of records created using the DTD within an EDI envelope, but it is also seen that records could be transmitted in other ways.

It is considered that the data dictionary of elements will be of use throughout the information flow from publisher to library.

8.2.3.6 Content

Within each media type, comprehensive coverage of bibliographic, trade and promotional data. The intention is to design a record that allows for different object levels associated with any line item (part works, bundles of works, multimedia packs etc). The overall structure is outlined in section 4.2.

8.2.3.7 Rules for formulation of content

Typically publishers do not follow cataloguing rules. This issue is not addressed.

¹ Martin, David. Data elements for an EDI 'Book Product Information' message. British National Bibliography Research fund Report 75. Published for BNB by BIC, 1995.

8.2.3.8 Extensibility

The DTD is being designed explicitly for extensibility across all publication types, and to provide for unforeseen requirements. Will be backed up by an organisation which represents all interested parties (publishers, wholesalers and retailers, libraries) and which can support development, dissemination and control.

8.2.4 EDIFACT EDI formats

In this section we deal with EDIFACT formats in a generalised way, i.e. talking about EDIFACT as an overall syntax rather than any specific message formats for metadata. We then discuss briefly the position with regard to carrying metadata in EDIFACT.

8.2.4.1 Community of use

EDIFACT is the accepted international standard messaging format for trading transactions in all industries. It is appropriate where bibliographic and other product information is communicated in the context of a trading relationship.

8.2.4.2 What is the level of maturity of the format?

EDIFACT is a mature and fully accepted standard although its practical application in the "book" trade has begun only in 1995/6. The EDIFACT syntax is maintained and developed by a world-wide process co-ordinated by a UN agency. The "book" trade has chosen to adopt an EDIFACT subset maintained by EAN International (EANCOM).

8.2.4.3 Controlling Agency

EDItEUR is the international group which interprets and extends EDIFACT message standards for "book" and serial applications.

8.2.4.4 How widespread is deployment

Use has begun in national and international trading between wholesalers and retail booksellers, and between library booksellers and libraries (see EU-funded EDILIBE project).

8.2.4.5 Technical considerations

Unlike SGML, the scope of the EDIFACT standard extends to a complete definition of a transmission envelope, identifying sender and receiver, and supported by a large number of "off-the-shelf" software packages and international VAN networks.

In theory a wide range of character sets is supported in the EDIFACT standard. In practice, they may not be implemented in current software.

8.2.4.6 Future path

EDIFACT continues to be developed to meet new requirements, but the standard is now pretty stable, and changes are generally upwards compatible.

8.2.5 Using EDIFACT to carry metadata

The EDILIBE project defined a method of carrying extensive bibliographic detail, based on UNIMARC, in certain message formats, to allow such detail to be sent between libraries and their suppliers as part of an approvals and ordering process. This approach has become part of current EDItEUR implementation guidelines.

EDItEUR is in the process of confirming a much simpler option, where basic bibliographic data are combined with price and availability information for the continuous updating of product databases used in trading. This may, and probably will, be a subset of one of the same messages which can be used for more extensive bibliographic data.

It can be assumed that both of these two options will continue to be supported, maintained and developed by EDItEUR, within the limits of existing EDIFACT syntax. A third option may be available soon, to carry within an EDIFACT envelope an "object" which is not itself in EDIFACT syntax. This could, for example, be a MARC record or an SGML document or another metadata format.

This last approach may be attractive where the parties to the exchange of information are in a trading relationship for which they are already equipped to use EDIFACT messaging. The extent of implementation needs to be considered, and whether these are the parties who will be involved in information flow from publishers to libraries.

It is unlikely that a "native" EDIFACT message will be developed to carry the full range of metadata which might be included in an SGML document, but it is still unclear exactly where a line should be drawn.

8.3 TEI, EAD and CIMI

TEI, EAD and CIMI will be dealt with in a separate study.

9. CATALOGUING ELECTRONIC PUBLICATIONS WITH MARC

9.1 Background

The various MARC formats are going through the process of being updated to enable cataloguing of electronic publications, in particular on-line publications. MARC format has the unique value for integrating metadata describing electronic resources into existing legacy systems. If libraries wish to integrate metadata into their existing systems, and use existing software (albeit with some updating to deal with new fields) then MARC offers a solution.

Most work has been done on adapting the USMARC format for the cataloguing items accessible through the Internet. OCLC's Intercat project has served as a test bed for the cataloguing of network resources, and as a means to introduce and verify new fields and fine tune as required. Over 200 libraries are participating in this project, the majority of them academic (60%) and nearly all of them situated in the US. There are now over 7,000 records in the Intercat database.

A guide has been produced as an output of the project:

- Olson, Nancy B. editor. Cataloguing Internet resources: a manual and practical guide. OCLC, 1995. Available at <URL: ftp://ftp.rsch.oclc.org/pub/Internet_cataloguing_project/Manual.txt>

Note that this guide is designed for OCLC users and examples and guidelines follow OCLC MARC rather than USMARC. In most circumstances there is little difference between the two formats.

The ISBD manual covering electronic resources is now awaiting final publication (after consultation). Following issue of the new cataloguing manual ISBD(ER), amendments may be required to UKMARC and UNIMARC.

9.2 UKMARC

At present the convergence programme between UKMARC and USMARC is underway. The policy is to adopt USMARC usage for the relevant fields relating to electronic publications and the proposed changes to UKMARC are going through the consultation procedure. The process involves dealing with particular sections of the format one at a time (1XX fields, 2XX fields etc). As part of this process 'new' USMARC fields for cataloguing electronic publications are being proposed. Because of the policy to address changes to UKMARC by sets of tag numbers, there has not been a single proposal for all the relevant fields, rather a gradual introduction of fields depending on the numbering of the tags. This can cause confusion in trying to identify all fields appropriate to cataloguing electronic resources.

It is worth noting that few UK libraries are currently cataloguing this type of material so there has been limited input experience of dealing with electronic documents. This is unlike the position with USMARC where the Intercat project gave the community experience in creating records, retrieving information using the records, and adapting current library management systems to handle the records.

9.2.1 *Those fields already accepted into the UKMARC manual:*

542 Mode of use note (machine-readable data files)

Note this is tag 538 in USMARC. A separate local access field is available in tag 856.

Examples

1. 542.00 \$a Mode of use: On-line video or teletype terminal or with a small dedicated computer (e.g. PDP8).
2. 542.00 \$a System requirements: 386SX processor; 6MB RAM; Windows 3.0; 7MB free space on hard disk.

9.2.2 Those fields already approved to be added to UK manual in an update:

856 Electronic Location and Access

The 856 field contains the information required to locate and retrieve an electronic item. The information identifies the electronic location containing the item or from which it is available. It also identifies the access method by which items can be retrieved. The information contained in this field should be sufficient to allow for the electronic transfer of a file, subscription to an electronic journal, or logon to an electronic resource.

Guidelines have been published by the Library of Congress on use of this field in USMARC:

- Guidelines for the use of field 856. Prepared by the Network Development and MARC Standards Office. Revised March 1996.

The 856 indicators and subfields can be used to indicate the access methods e.g.

Email, FTP, Remote login (Telnet), Dial-up, http, wais etc.

Subfield \$u contains the Uniform Resource Locator (URL), which provides electronic access data in a standard syntax. Field 856 is structured to allow the creation of a URL from the concatenation of other separate field 856 subfields. Subfield \$u may be used instead of those separate subfields or in addition to them. Subfield \$u may be repeated if the other information in the field applies.

This field also allows for inclusion of:

- access number associated with a host. It can contain the Internet Protocol (IP) numeric address if the item is an Internet resource, or a telephone number if dial-up access is provided through a telephone line
- size of file
- hours of availability but if the record is for a system or service hours are recorded in field 307(Hours, etc.)
- Subfield \$y contains the access method when the first indicator position contains value 7 (Method specified in subfield \$y). This subfield may include access methods other than the three main TCP/IP protocols specified in the first indicator. The data in this subfield corresponds with the access schemes specified in Uniform Resource Locators (URL). The Internet Assigned Numbers Authority (IANA) maintains a registry of URL schemes and defines the syntax and use of new schemes; the British Library will include an authoritative list based on that standard in *USMARC Code List for Relaters, Sources, Description Conventions*.

Examples

1. 856 10\$uftp://path.net/pub/docs/urn2urc.ps
2. 856 70\$uhttp://lcweb.loc.gov/catdir/semdigdocs/seminar.html\$yhttp
3. 856 30\$alocis.loc.gov\$b140.147.254.3\$mlcon-line@loc.gov\$t3270\$tline mode(e.g., vt100)\$vM-F 6:00 a.m.-21:30 p.m. USA EST, Sat. 8:30-17:00 USA EST, Sun. 13:00-17:00 USA EST.
4. 856 10\$awuarchive.wustl.edu\$dmirrors/info-mac/util\$fcolor-system-icons.hqx\$s16874 bytes
856 00\$akeptvm.bitnet\$facadlist file1\$s34,989 bytes\$facadlist file2\$s32,876 bytes\$facadlist file3\$s23987 bytes.

9.2.3 Fields contained in British Library December 96 consultative document:**258 Computer File Characteristics (NR)**

This field is used to record characteristics pertaining to a computer file. It may contain information about the type of file (e.g. Computer programs), the number of records, statements, etc. (e.g. 1250 records, 5076 bytes).

Examples

1. 258 00\$aComputer data (2 files: 876,000, 775,000 records).

307 Hours, etc. (NR)

This field contains chronological information identifying the days and/or times an item is available or accessible. It is used primarily in records for electronic resources.

Examples

1. 307 00\$aM-F, 9AM-10PM
2. 307 00\$aTu-F, 10-6; Sa, 1-5, USA PST.

753 System Details Access to Computer Files

This field contains information relating to the type of machine, operating system, and/or programming language used with computer files and accompanying material. This kind of added entry is assigned to give access to the bibliographic record which otherwise would not be possible and to facilitate the capability of selecting and arranging records for production of printed indexes.

Examples

1. 753 00\$aIBM PC\$bPascal\$cDOS 1.1
2. 753 00\$aCompaq\$bBasic\$cDOS 3.2
3. 753 00\$aApple II\$cDOS 3.3.

537 Type of Computer File or Data Note

In the consultative document it is suggested that this be expanded to include notes held in USMARC 516.

Examples

1. 537 00\$aSPSS 5.2
2. 537 00\$aText(Law reports)
3. 537 00\$aComputer program.

9.3 UNIMARC

9.3.1 *Background*

Most fields to do with electronic cataloguing are marked as provisional in UNIMARC - awaiting ISBD(ER) to be published. The exception is US/UKMARC 856, which will be considered at the next Permanent UNIMARC Committee (PUC) meeting in March 1997.

A draft version of UNIMARC Guideline 3 for Computer Files was issued in June 1995 and is now the January 1997 consultative document. (Note the renumbering of Guideline 3, as Older monographic publication guidelines was published with that number). These guidelines result from meetings of the IFLA Permanent UNIMARC Committee and the requirements of the International Standard Bibliographic Description for Computer Files, ISBD(CF). A new draft of this Guideline is expected soon. The current version can be made available, but the published version will take account of ISBD(ER), with a timetable dependent on the publication date of the latter. (ISBD(ER) in a later revision of ISBD(CF)).

The Guidelines for Computer Files have been formulated with off-line products in mind i.e. CD-ROMs, diskettes. No special fields such as URLs are specified for metadata specific for networked resources. The Guidelines are still being developed to be better suited for on-line materials as well.

9.3.2 *Fields used for cataloguing electronic resources*

The Guideline 3 specifies the use of existing fields for the description of computer files, but in addition other data elements from UNIMARC may be used in a record for a computer file. The probable need for additional fields or content designators and for redefinition of existing fields in the near future is acknowledged.

The following fields are specified in the guidelines:

- title (field 200): Title as it appears on container, box, opening screen, formal title screen, first display of information, header of the file etc.
- parallel title (field 510): Title in another language appearing on the computer file
- author(s) (fields 200\$f \$g and/or 700, 701, 710, 711): Authors, programmers of the computer files as listed on the computer file
- author affiliation(s) (fields 700\$p, 701\$p, 710\$p, 711\$p): Institutional affiliations of the authors, programmers at the time the computer files were written or programmed
- edition statement (field 205): Any word or phrase indicating that the information was available previously in a different form
- publication, distribution (field 210)
- physical description of the computer file (fields 215, 230): To be omitted for remotely accessed computer files, because there is no physical item
- accompanying materials (fields 215, 307): User handbooks
- series (fields 225, 410)
- availability information (fields 345, 010, 011): Price units, stock number, agency for ordering a copy of the computer files.

Apart from the above mentioned fields, some of the (extra) information should be put in different fields of the Note block (3XX). This concerns the following data:

- type of computer file (field 336)
- technical details of computer file (field 337)
- notes pertaining to title and statement of responsibility (field 304)
- notes pertaining to edition: (Licensed by...) (field 305)
- notes pertaining to publication, distribution (Shareware, etc.) (field 306)
- notes pertaining to series (field 308)
- notes pertaining to availability (field 310)
- contents notes

- users/Intended audience note.

The 1XX block provides fields for:

- coded data
- qualifying data
- language of computer file
- target audience
- publication date
- country of publication or production
- coded data relating to computer files: program, representational, textual.

In Guideline 3 no special field is provided yet for information pertaining to location. USMARC 856 is being examined to see if it can be adopted for UNIMARC.

Field 135 is the provisional Coded Data Field for Computer Files. For type of computer file and technical details fields 336 and 337 in the Notes block are defined.

- in field 135, a one-character code indicates the type of data file:
 - a = numeric
 - b = computer program(s)
 - c = representational (pictorial or graphic information)
 - d = text
 - u = unknown
 - v = combination
 - z = other
- type of computer file (field 336): contains information characterising the type of computer file. In addition to a general descriptor (e.g. text, computer program, numeric), more specific information, such as the form or genre of textual material (e.g. biography, dictionaries, indexes) may be recorded in this field.
- technical details note (field 337): This field is used to record technical information about a computer file, such as the presence or absence of certain kinds of codes or the physical characteristics of the file (e.g. recording densities, parity, or blocking factors). For software, data such as the software programming language, the number of source program statements, computer requirements (e.g. computer manufacturer and model, operating system, or memory requirements), and peripheral requirements (e.g. number of tape drives, number of disk or drum units, number of terminals, or other peripheral devices, support software or related equipment) can be recorded.

9.3.2.1 Host administrative data:

No fields are specified for information pertaining to the host. USMARC practice may be adopted for UNIMARC.

9.3.2.2 Administrative metadata

There are no fields for record review date and creation date.

9.3.2.3 Provenance/source

Availability information is included in fields 345 (Acquisition information note), 010 (ISBN) and 011 (ISSN). Further notes pertaining to availability go in field 310 (Notes pertaining to binding and availability).

9.3.2.4 Terms of availability/copyright

The relevant USMARC fields are being examined for this purpose.

9.3.2.5 Rules for the construction of these elements

Field 801 (Originating Source), subfield \$g, contains an abbreviation for the cataloguing code used for bibliographic description and access. The Manual gives a list of the accepted codes in an appendix.

Other codes may be registered with the IFLA UBCIM Programme.

9.3.2.6 Multi-lingual issues

Character positions 26-29 and 30-33 of field 100 subfield \$a are used to designate the default and additional graphic character sets used in the record. Sets approved for use with UNIMARC are:

- ISO 646 (IRV), Basic Latin set
- ISO 5426-1980, Extended Latin set
- ISO Registration #37, Basic Cyrillic set
- ISO DIS 5427, Extended Cyrillic set
- ISO 5428-1980, Greek set
- ISO 6438-1983, African coded character set.

10. MAPPING METADATA FORMATS TO MARC

10.1 Mapping of IAFA Templates to OCLC MARC

The following are proposed mappings tabulated by Michael Day at UKOLN. Various other useful mappings have been proposed elsewhere such as Dublin Core to EAD and GILS (Eric Miller. Dublin Core element set crosswalk. available at <URL:<http://www.oclc.org:5046/~emiller/DC/crosswalk.html>>).

Table 1. Mapping from IAFA templates to OCLC MARC

IAFA Template Handle:	MARC File:	
Category	655 518	Index Term - Genre/Form, or Type of Computer File or Data Note
Title	245\$a	Title Statement
URI-v	856	Electronic Location and Access
Short-Title	246	Varying Form of Title
Alternative-Title	246	Varying Form of Title
Author - (USER*)	100 110 700 710 245\$c	Main Entry - Personal Name, or Main Entry - Corporate Name, or Added Entry - Personal Name, or Added Entry - Personal Name, or Title Statement
Admin - (USER)*	710	Added Entry - Corporate Name
Source	500	General Note
Requirements	538	System Details Note
Citation	524	Preferred Citation of Described Materials Note
Publication-Status	500	General Note
Publisher - (ORGANISATION*)	260\$b	Publication, Distribution, etc.
Copyright	500	General Note
Creation-Date	260\$c	Publication, Distribution, etc.
Discussion	500	General Note
Keywords	653	Index Term - Uncontrolled
Version-v*	250	Edition Statement
Format-v*	538	System Details Note
Size- v*	256	Computer File Characteristics
Language- v*	Lang:	
Character-Set- v*	500	General Note
ISBN	020	ISBN
ISSN	022	ISSN
Last-Revision-Date- v*	260\$c	Publication, Distribution, etc.
Subject-Descriptor-Scheme	See below	
Subject-Descriptor- v*	050 080 082 084 090 092	Library of Congress Call Number UDC Number Dewey Decimal Call Number Other Call Number Locally Assigned LC Locally Assigned Dewey

IAFA Template Handle:	MARC File:	
	098 6XX	Other Classification Schemes Subject Added Entries
To-Be-Reviewed-Date		No equivalent
Record-Last-Verified-Email		No equivalent
Record-Last-Verified-Date		No equivalent
Comments		No equivalent
Destination		No equivalent

10.2 Mapping of Dublin Core to USMARC

This mapping is based on LC Discussion Paper 86 (MARBI, 1995).

Table 2 mapping Dublin Core to USMARC

Dublin Core element	USMARC
Subject	653 Index Term -- Uncontrolled, or 650 Subject Added Entry -- Topical Term.
Title	245 Title Statement.
Author	100 Main Entry -- Personal Name, or 110 Main Entry -- Corporate Name, or 700 Added Entry -- Personal Name, or 710 Added Entry -- Corporate Name.
Publisher	260\$b Name of Publisher, Distributor, etc.
OtherAgent	700 Added Entry -- Personal Name, or 710 Added Entry -- Corporate Name.
Date	260\$c Date of Publication, Distribution, etc.
ObjectType	Leader/06 Type of Record.
Form	538 System Details Note.
Identifier	010 LC Control Number, or 020 ISBN, or 022 ISSN, or 024 Other Standard Identifier, or 856\$u Uniform Resource Locator.
Relation	772 Parent Record Entry, or 773 Host Item Entry, or 775 Other Edition Entry, or 776 Additional Physical Form Entry, or 780 Preceding Entry, or 785 Succeeding Entry, or 787 Non-specific Relationship Entry.
Source	786 Data Source Entry, or 776 Additional Form Entry.
Language	041 Language Code, or 546 Language Note.
co	Spatial: 034 Coded Cartographic Mathematical Data, or 255 Cartographic Mathematical Data. Temporal: 045 Time Period of Content, or 513 Type of Report and Period Covered Note.

10.3 Mapping of Dublin Core to PICA Format

At the moment Koninklijke Bibliotheek of the Netherlands as part of the WebDOC project are working on a conversion model to map the minimum requirements of metadata they would like to receive from publishers to the extended PICA format.

A working group of PICA participants from Germany and The Netherlands is composing a checklist of metadata (inspired by the 'Dublin Core'). The metadata of this list which could be embedded in a publication are to be automatically converted to the contents of tags in the PICA format. For the moment the proposal for the list is:

Table 3 : Mapping Dublin Core to PICA format

Labels	Structure	PICA+
Author	<Name>, <First name> <(Addition)>	028A\$d<Firstname>\$a<Name \$y<Addition>
Corporation as author	Free text	029F/01 \$a
Next corporation	Free text	029F/02\$a
Title	Free text	021A \$a
Place of publication	Free text	033A \$p
Publisher or host	Free text	033A \$n
Next place of publication	Free text	033B/01 \$p
Next publisher or host	Free text	033B/01 \$n
Date, year	dd-mm-yyyy	011@ \$ayyyy \$ndd_mm_yyyy [converted from table]
Language	In English: free	010@ \$axx [converted from table]
Next language	In English: free	010@ \$bxx [converted from table]
Edition/update	Free text	032@ @a
Form	Multiple choice	009/P03 \$0
URL	Structured	009/P03 \$a
Subject keywords	In English: free	144Z \$a
Next subject keywords	In English: free	144Z \$a
Abstract	Free text in English	047I \$a

Note that this a proposal of KB to WebDOC and has not yet (October 1996) been discussed in the working group.

11. CONCLUSION

The format of metadata used in BIBLINK depends on the requirements we are trying to fulfil. There seem to be two strands that can be followed and both have their merits:

- the provision of timely information in the form of brief details in advance of publication (or at least at the time of publication)
- the provision of detailed 'exhaustive' information by utilising the full information held by the publisher in their own internal databases.

As Mark Bide argues², in order to minimise waste in the supply chain the aim should be to identify the point at which the earliest possible quality-controlled and standardised key stroke can be captured; and this implies capture of the record keyed in by the publisher (or author). This argument seems valid whatever the format of metadata (simple or complex).

In addition, as part of the consensus building process there are other outstanding questions that need to be answered. These are summarised as follows:

Recommendation

Project participants need to agree a clearer definition of

- **level of detail required as regards content of metadata (simple CIP record or a richer record)**
- **what object level national libraries will want to describe (the level of granularity of resource that will be described)**
- **to what extent content of metadata can be covered by rules for formulation of content (i.e. cataloguing rules)**
- **reaffirm whether within the project the aim of the project is to catalogue electronic publications from a few well established publishers or to address improving bibliographic control of the wider world of web publishing.**

The choice of metadata format cannot be divorced from the business model. Creation of records involves investment, and the level of investment increases with the complexity of the record structure. We need to consider the likely level of investment in the creation of metadata for electronic resources, who will be making this investment and how this will affect the metadata format. If the intention is for national libraries to create detailed MARC records how many electronic publications will be selected for this detailed treatment? Are we considering selecting a small number of stable electronic publications for description, or is there an ambition to provide some form of bibliographic control to the wider world of web publications? It seems the choice of metadata standard in this context cannot be divorced from the commercial environment. If publishers are to provide detailed information to national libraries then they may well wish to retain rights in the final record which appears in the national bibliography. Certainly some publishers would be unwilling to include descriptions and TOCs to be included without some commercial arrangement.

² Bide, Mark (on behalf of Book Industry Communication) Electronic Tables of Contents (EToCs) for serials: standards for structure & transmission. A research study for the BNB Research Fund. Fourth Draft 1994.

Recommendation

We need to make explicit assumptions about the business model for the provision and exploitation of metadata.

If we intend to use a standard that exploits the 'single record' produced by the publisher then the choice must be SGML syntax. This is the syntax with which publishers are familiar and it will allow publishers to re-use the header information they are already creating. The choice of SGML also means that web based documents will be compliant in so far as HTML is compliant to SGML.

From experience we see that publishers cannot agree on a single DTD for article headers, so it would seem they are even less likely to agree on a DTD for all non-serial electronic publications.

In order to encompass the diverse range of publishers and material involved in the process, the use of a minimum set of data is attractive. If we are looking towards use of a core set then the Dublin Core element set is an obvious choice. There is international involvement in the consensus building, and project participants could influence the development of the format. It is a format that small publishers and web publishers could use without incurring significant overhead.

More detailed information could be provided using the data element set defined by the BIC SGML DTD for non-serial publications now under development, and the SSSH DTD for article headers. It is planned to have a minimum set defined in the non-serial DTD and a minimum set is already defined in SSSH. It would seem useful if these minimum sets could encompass at least the Dublin Core elements. If this was the case then the more complex record could be mapped to Dublin Core, either to create a separate Dublin Core record for transmission or as a means to allow interoperability during transactions involving the record. Other services supplying detailed records (e.g. archives using TEI headers) could also map more complex information to Dublin Core.

Recommendation

It seems one metadata format may not be sufficient for the diverse body of publishers described in the scoping document. It would be more realistic to consider two formats to allow for the creation of a brief record and a more complex record.

It might be possible for more detailed records made available from the publisher (when such records exist) to be supplied as a separate physical record in addition to Dublin Core set. It would seem an attractive opportunity to implement the Warwick Framework architecture to enable the simple Dublin Core record to be packaged together with the more detailed SGML header. The Warwick Framework might also offer possibilities for defining terms and conditions which could be applied to use of different metadata (e.g. to charge for use of Table of Contents).

Recommendation

Consider use of Dublin Core as a minimum element set. Consider use of BIC non-serial DTD and SSSH for more complex records. Consider implementation of Warwick Framework to package more complex SGML records with Dublin Core records.

The provision of bibliographic services in the context of electronic publications is an unstable environment. The nature of the resources themselves are changing, the metadata formats are not mature, the commercial and service model is uncertain. As regards use of metadata it seems clear there is no one solution. New players in the 'information' world outside the traditional library world (e.g. browser manufacturers, commercial Internet search services) will influence future development of formats, just as technology affected development of MARC. In this environment it seems mapping and interoperability between formats will remain an issue for the foreseeable future.

In the context of BIBLINK we must recognise that none of the formats we might use are fully agreed. This includes UKMARC and UNIMARC both of which are being updated to deal with electronic publications. UKMARC is in additional flux due to the convergence program. In this situation any solution must take into account the need for changes to particular fields and data elements and ensure these can be accommodated.

Recommendation

We accept that all metadata formats in this area are unstable. We need to define what level of maturity

and stability are required in our format(s) of choice. At that stage we may wish to influence the development of the format(s).

ANNEX A. PRODUCT INFORMATION FLOWS IN THE PUBLICATIONS SUPPLY CHAIN**A.1 Background**

The purpose of this document is to outline the major flows of product information from publishers through various other agencies to libraries as they currently exist, and to identify the most widely used data standards and communication media. The present draft is based almost entirely on knowledge of the UK situation. Other strands may need to be incorporated to cover patterns which have been adopted elsewhere; and there are probably some features which are at present unique to the UK. For the purposes of product or bibliographic information (metadata) transfer from publishers through the supply chain to libraries, it is necessary to regard electronic publications as part of a continuum which begins with conventional print; for while network publishing has some entirely new characteristics which cannot be shared with other forms, physical carriers of digital material such as diskette and CD-ROM can be and often are distributed in exactly the same way and through exactly the same channels as printed books, audio and video cassettes, and other media - frequently as parts of a hybrid package.

A.2 Organisations involved in information transmission

The organisations involved in the transmission of metadata related to serial and non-serial publications have been categorised as falling into eight types under four major headings:

1. Publishers (including distributors handling the output of a number of different and independent publishers; these may be a source of some product information on behalf of their clients).
UK examples of publishers substantially involved in a mix of print and non-print publishing and/or with relatively well-developed internal information publishers substantially involved in a mix of print and non-print publishing and/or with relatively well-developed internal information systems: Oxford University Press, Institute of Physics, Dorling Kindersley, Reed Group (both academic and general trade publishing)
2. Information services (ie compilers and suppliers of metadata independent of the supply of actual publications)

2.1 National library bibliographic agencies.

UK example: the British Library National Bibliographic Service.

2.2 Trade bibliographic agencies.

UK examples: Whitaker, Book Data, Bibliographic Data Services.

2.3 Abstracting and indexing services, typically subject-based.

UK example: INSPEC (The Institution of Electrical Engineers).

2.4 Serial contents databases, not subject-based.

These may be associated with a journal subscription agent or national library document supply service, but have been treated as a separate category for the purposes of this analysis. UK examples:

Blackwells, British Library Document Supply Centre.

3. Suppliers (ie suppliers of actual publications who also supply metadata)**3.1 Booksellers and library suppliers.**

UK examples of booksellers particularly active as

(a) users of externally-supplied metadata in their own business: W H Smith, Dillons, Heffers, or

(b) onward suppliers of metadata to library customers: Askews, Dawsons, Holt Jackson, JMLS/Blackwells etc.

3.2 Journal subscription agents.

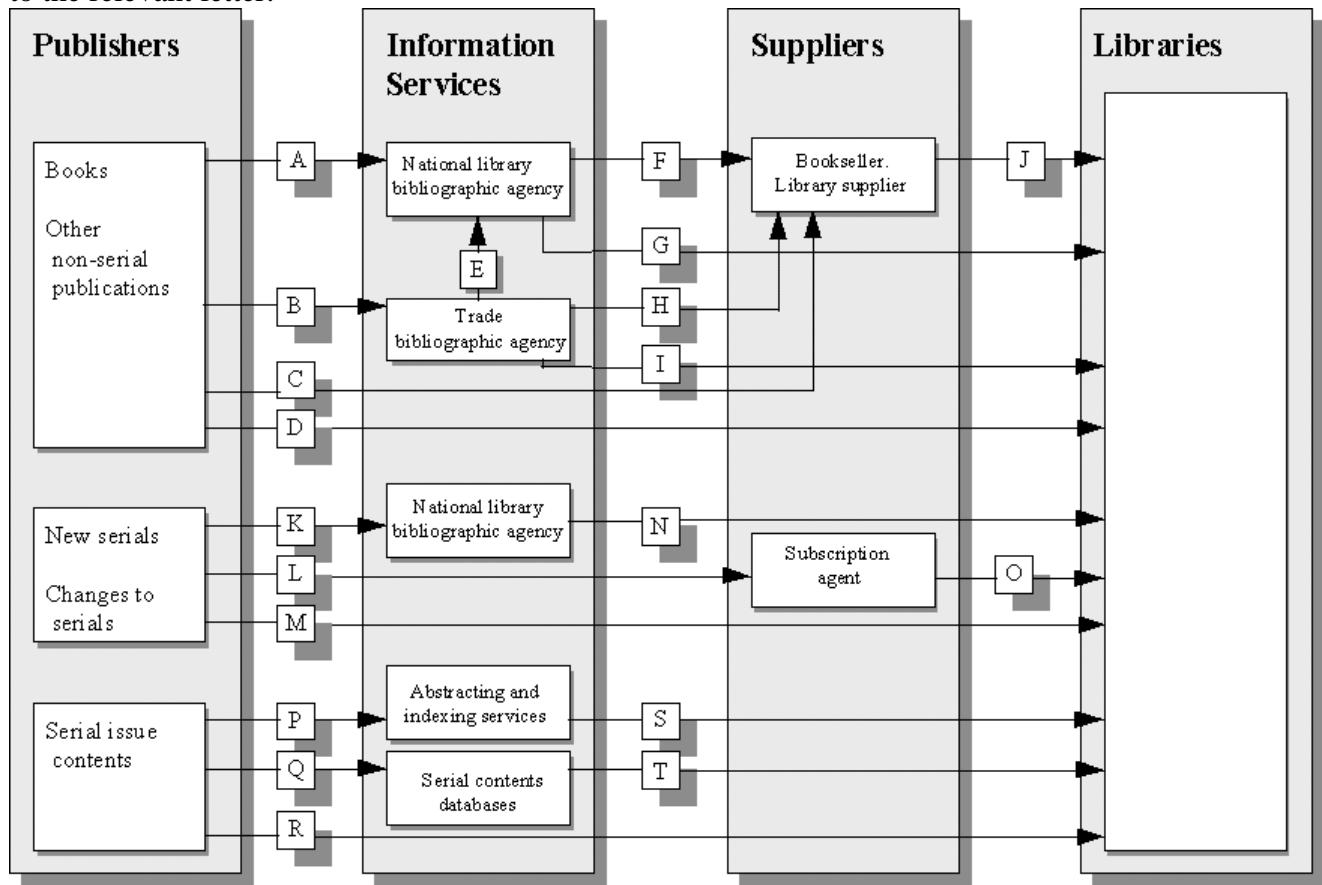
The major agents are international: Blackwell Group, Dawsons/Faxon, Swets, EBSCO etc.

4. Libraries

It has not been considered necessary to identify library co-operatives and system suppliers (eg BLCMP, SLS, CURL) as a separate category, although they play a definite part in facilitating access to metadata.

A.3 Metadata information flow diagram

The accompanying diagram charts the main flows of metadata between the types of organisation identified above. Each flow line is labelled with a letter (A to T), and the commentary below is keyed to the relevant letter:



A.3.1 Books and other non-serial publications

This section of the diagram includes print and non-print (eg CD-ROM, video, audio) where it is published and distributed within the book supply chain. It is worth noting that Book Industry Communication (BIC) has developed a set of guidelines for the content of publishers' in-house bibliographic databases [1]. Although these do not define a transmission format as such, they are intended to support record supply to other agencies. The guidelines have been applied by some UK publishers.

- A Publisher to national library bibliographic agency. Traditionally this was based on paper forms prescribed by the agency, and was limited to pre-publication (CIP) data, with a definitive record being based on the legal deposit copy of the work; and it was (and may still be) restricted to printed publications. No standard is known to be in use for the electronic transfer of metadata from publishers to national libraries.
- B Publisher to trade bibliographic agency. Two important differences between this line and line A are that advance information may be communicated to a trade agency very much further ahead of expected publication (eg up to two years for some titles) and that updates will typically continue after publication so that the trade agency can show current prices and availability. As yet, no standard is widely used for electronic transfer at this level, but two national EDI formats exist: the BIC/ANA TRADACOMS Price & Availability Update message [2], which Book Data and others are bringing into use in the UK, but which is not designed to carry full bibliographic data; and the US BISAC X12 832 Price/Sales Catalog message [3], which will support a full bibliographic record, including a semi-structured table of contents.
- C Publisher to bookseller/library supplier. Most communication from publishers to booksellers remains on paper (eg new title advance information sheets, current stocklists), but a number of publishers offer electronic information direct to booksellers, generally in a proprietary format, eg as a CD-ROM catalogue.
- D Publisher to library. Again most communication has been on paper. Publishers' CD-ROM catalogues are also distributed to libraries. A recent development is the rapid growth of publishers' web sites, for which libraries are presumably a significant part of the target audience.
- E Trade bibliographic agency to national library bibliographic agency. In the UK the British Library sub-contracts the CIP programme by competitive tender, currently to Bibliographic Data Services. Information supply is in UKMARC format [4], and the principal medium is understood to be Internet file transfer.
- F National library bibliographic agency to bookseller or library supplier. Some library suppliers incorporate MARC records into their own databases, and/or redistribute them to libraries. The format is MARC, and supply (in the UK) may be on magnetic media, CD-ROM or by Internet file transfer.
- G National library bibliographic agency to library. MARC format; the dominant medium for supply to individual libraries is either CD-ROM or on-line, but magnetic media and Internet file transfer are also possible.
- H Trade bibliographic agency to bookseller or library supplier. MARC format on magnetic media or by Internet file transfer is likely to apply where the bookseller requires MARC for redistribution to a library clientele. Proprietary or tailored formats apply elsewhere. The BIC EDI update format [2] is coming into use. CD-ROM is the dominant medium for smaller booksellers. On-line is used scarcely or not at all. Trade bibliographic agency to library. CD-ROM, with MARC download, is probably the dominant medium, together with on-line access through library host systems.
- J Bookseller or library supplier to library. Predominantly MARC format, in a variety of media; but note also the recent development of suppliers' CD-ROMs, replacing approval collections, and in some cases offering download of order records in a TRADACOMS or EDIFACT EDI order format to be passed to the library acquisitions system.

A.3.2 New serials and changes to serials

- K Publisher to national library bibliographic agency. Notification of new serial titles and substantial changes are sent to the agency as national centre for ISDS, responsible for the assignment of ISSN, and are incorporated into the national bibliography. Electronic transfer does not apply at this level.
- L Publisher to subscription agent. There is necessarily a continuous flow of product information from journal publishers to subscription agents. Some electronic transfer takes place in proprietary formats.
- M Publisher to library. Essentially the same as D above. Publishers' CD-ROM catalogues and web sites are likely to cover serial and non-serial publications.
- N National library bibliographic agency to library.
- O Subscription agent to library.
- A.3.3 Serial issue contents**
- P Publisher to abstracting and indexing service. Traditionally A&I services have worked from the printed copy of the journal issue, but some are receiving electronic copy from a limited but growing number of publishers. Available standards are SGML-based: MAJOUR [5] and SSSH [6].
- Q Publisher to serial contents database. Essentially the same as P above.
- R Publisher to library. There is some direct supply of electronic tables of contents from large publishers to individual libraries, by Internet file transfer or web site.
- S Abstracting and indexing service to library. Dominant media are on-line and CD-ROM, with proprietary formats?
- T Serial contents database to library. Dominant media are on-line and CD-ROM, with proprietary formats?

A.4 Existing standards referred to in the text

- [1] BIC Manual on Publishers' Bibliographic Databases, draft, 1994 (unpublished, but available on request from Book Industry Communication).
- [2] Book Trade Price & Availability Updates File, TRADACOMS File Format 108, Version 1, London, Book Industry Communication/Article Number Association, July 1993.
- [3] BISAC X12 832 Price/Sales Catalog, New York, Book Industry Study Group, 1996
- [4] UKMARC Manual
- [5] MAJOUR: Modular Application for Journals, European Workgroup for SGML, 1991
- [6] SSSH: Simplified SGML for Serials Headers, London, Book Industry Communication/PIRA International, 1996

A.5 Standards under development

The following standards are the subject of development work which is currently in hand in national and international groups representing the book and serials sectors:

- SGML formats for book and serial product information
- EDIFACT formats for book and serial product information

**ANNEX B. NATIONAL LIBRARIES METADATA REQUIREMENTS: QUESTIONNAIRE
REPLIES****B.1 Bibliotheque nationale de France****1. Please indicate the cataloguing system you use.**

BnF is building a new integrated information system that will be operational in 1998. We decided to provide you with information available for existing systems. They will be operational until 1998/1999. BnF is cataloguing only OFF-LINE electronic documents.

As BnF is the French ISSN centre we have started attribution of ISSN numbers to electronic journals (see scoping document). It is transmitted to the international ISSN data base but NOT included in the national bibliography.

A distinction should be made between the format and the cataloguing system.

Format used : INTERMARC

Exchange format : UNIMARC

Cataloguing rules: AFNOR (National French Standardisation body) standards for cataloguing rules.

Cataloguing rules for electronic documents :

We use the French AFNOR standard : Z 44-082 "catalogage des documents electroniques". It is based on ISBD (CF) version 1.

It applies to both on-line and off-line documents.

Databases :

BN-OPALE 1 800 000 records

Electronic documents (for databases) are recorded in this database.

System :

BN-OPALE GEAC

BN-OPALINE

Bibliographic database for specialised departments

Electronic documents (multimedia electronic documents) are recorded in this database.

System :

BULL database system

2. What data elements do you record for an electronic publication in addition to those you record for a traditional publications (see BnF sample records attached)?

All the information below is both for MONOGRAPHs and SERIALS

For electronic documents, following ISBD (CF) we include :

- AREA 1 - after the proper title between [] we specify : [general material data = electronic document]
- AREA 3 - type and size of electronic document
- AREA 5 - technical description (= requested configuration)

3. Which MARC (or equivalent) fields do you use to hold the data element ?

INTERMARC FIELDS

Descriptive fields

AREA 1 = 245 \$d

AREA 3 = 257 \$a type \$b size

Fixed fields

Leader

position 7 : 2 values of position 7 : s (program) or t (data)

zone 008

position 23 : s (CD-ROM, CD-I)
 t (disquette)

Note field

AREA 5

field 336 : subfield \$w

field 337 : subfield \$k = requested configuration
 subfield \$a = technical information on material

For electronic documents **holdings** : in field 036 = legal deposit number.

We do not include for the moment any data concerning citation, communication...

In future, we will be able to create a link between the record for the electronic version and records for the same document on another different support (electronic ---- print).

4. What data would you like to include in the record that you cannot find or have difficulty finding from the publications ?

For electronic documents the cataloguing is made from the document itself. This means that we "open" the document. In case of a CD-ROM we do install it on a PC to open it and thus create the record from the content of the disk (not just looking at the commercial printed Technical installation).

We have added complementary information for technical information: we add the type of material used for installation in BnF. This information may be different from the technical / marketing information mentioned by the publisher under the section "configuration requested".

We also add the time of installation.

This information will be later on important for access to the documents by the end-user. If you need half an hour to load a document this has to be taken into consideration for the access service.

4.2 Source of description

The information needed to create the bibliographic record is not gathered in one part of the document as it is in a cover page for a book. You need to browse carefully the CD-ROM and collect the information needed at different places in the CD. It is time consuming. There is no structuring of bibliographic information. This could be requested from the publisher to help the cataloguing.

4.3 Multiple author/ multiple functions

For a multi-media electronic document there is a multiplicity of "authors" based on different functions. The rule is to limit to three authors for the same function. But for a CD-ROM you can have more than 10 different functions : script developer, infographist, designer, title manager, music....This increases the workload for bibliographic description of "AUTHORS".

4.4 Information on terms and conditions should also be included.

For our own digitised collection, we mention information on terms and conditions in the holdings zone.

4.5 We need precise date of publication.

6. What library services will use metadata for electronic publications ?

1. National Bibliographic Agency both for legal deposit and National Bibliography.
2. Specific bibliographic by-products for electronic documents
3. Services in charge of access to the collections.

7. What metadata elements do you consider will be required for electronic publications ?

If we consider the Dublin Core metadata element the following elements could be re-used for bibliographic description of electronic documents :

see also question 4.

- Author
- Other Agent (see remark question 4)
- Publisher
- Title
- Date
- Object type
- Form
- Identifier
- Relation
- Source
- Language.

Subject : this field raises the issue of the list of references (thesauri) used by the producer of this information. But it can be useful for the cataloguer who will then adapt it to the indexing system used in the library.

If not available in Dublin Core the following information may also be useful :

- Mention of the Host if different from the publisher.
- Mention of the version of the document : Is it the final version or an updated version ?

- Mention of the availability of the document in terms of duration : for how long will an electronic document be available at the URL indicated ?

The question of long term availability of the document can also be discussed. As long as there is no national authority in charge of collecting on-line electronic documents a mention of responsibility for guaranteeing the access to the document in case of a change in the URL can be added. Who is responsible ? the author, the publisher, the host ?

8. Will the records for electronic publications need to be integrated in your existing systems ?

For off-line electronic publications there are already integrated in our system. For on-line, it will be the same. We will have a unique multimedia integrated system for the OPAC and national bibliography. This requirement is very important because we will adapt the existing system to new electronic documents but we will not completely change our system or our format in the next five years.

9. Will the metadata need to be manipulated by particular protocols ?

Yes, the catalogue will support Z39-50 protocol. We plan to give access possibilities to other libraries to the BnF national bibliography records via FTP.

10. Please add any further comments about your experiences in cataloguing electronic publications.

BnF has a solid experience for cataloguing off-line electronic publications. The on-line topic is very important to us and we will benefit from BIBLINK results before starting a national project. The French national bibliography is based on the legislation for legal deposit which includes off-line electronic publications. The problem is to include progressively on-line publications but then we come back to the limits. What do we include : electronic journals, home pages (on which criteria ?)...The evolution towards cataloguing "articles" is also an important issue as we do not include articles in the national bibliography.

B.2 Royal Library response

1. Please indicate the cataloguing system you use (for example, UNIMARC), and which cataloguing rules you apply.different authority

The Koninklijke Bibliotheek, National Library of The Netherlands, is using the cataloguing system of PICA in Leiden. PICA, the Library Automation Centre in the Netherlands, maintains a database with more than 11.000.000 title records accompanied by records from files: thesauri for personal and corporate authors, common and local subjects, printers and editors, etc. This database is in common use by the KB (i.e. for the Dutch National Bibliography), most university libraries, large public libraries and NBLC, i.e. a company supporting all public libraries.

The Shared Automated Cataloguing system (GGC) is used for exporting data to different OPAC's. The Dutch automated Union Catalogue uses this database for Inter Library Loan. For the German market PICA developed databases with the same structure. On-line input to the database is possible by using the PICA cataloguing format. The PICA format as used in the database is like MARC formats built with tags, fields and subfields. The format for the cataloguers is developed to be more user friendly. In this format for the cataloguers many subfields are indicated with the punctuation used in the ISBD rules. In this document PICA tags are mentioned in the internal PICA+ format. The 'cataloguing tags' are added between parentheses. The PICA format is specified to be able to display the data of title records according to the rules for cataloguing (ISBD).

The rules for the use of the PICA format for electronic publications are published in several documents. These are still draft and they are being developed further while we gain experience in cataloguing on-line and off-line electronic publications. Titles of relevant documents are:

- Richtlijnen voor het catalogiseren van Computer Files (Guidelines for cataloguing Computer Files)
- Richtlijnen voor het catalogiseren van on-line resources DN54/0895, Leiden 1995 (Guidelines for cataloguing on-line resources DN54/0895. HTML version by SURFnet Infoservices, The Hague 1995.
URL <http://www.konbib.nl/kb/is/gedocarchief/edocarchief/olr.en.htm>
- Examples of titles of On-line Resources can be found on:
<http://www.konbib.nl/kb/is/gedocarchief/edocarchief/titles.htm>

2. What data elements do you record for an electronic publication in addition to those you record for a traditional publication? Also, what data elements do you not record that you would use for a traditional publication? (Sample records may be attached.)

Answer to questions 2 and 3.

To avoid redundancy the answers to these two questions are combined. The KB is gaining experience now with cataloguing electronic on-line and off-line documents in the PICA cataloguing system. For the moment two formats are used in the system: a format for On-line Resources, especially used for WWW and a format in use since 1986 for off-line software products but nowadays also used for off-line publications like CD-ROM's, diskettes, CD-I's etc. PICA and KB are co-operating to develop a format for all electronic resources, both on-line and off-line. In the answers below it has been attempted to indicate what is now already fully implemented in the production line and what is in a phase of testing. Below is described the practice today in KB.

a) Off-line monograph (ie CD-ROM encyclopaedia) =

- 011@ (1100) \$a Year of publication; \$n more specific information; example: \$a1995 \$b01 Apr. 1995.
In electronic resources, more than in folio editions, the exact date of issue is important to distinguish updates.
- 016A (1101) \$a (proposed) Coded information: pos. 2: specific material designation: tape, disc, etc.; pos. 3: type computer file: data, program, etc.; pos. 4: additional information: numeric data, text, sound, etc.
- 039Q (4262) \$9 The PICA Production Number (PPN) of the record in the PICA database of the original printed version.
This tag realises links between the records of the electronic and the printed versions of the same publication.
- 048H (4251) \$a System requirements: for off-line: kind of computer, processor, memory, operating system, monitor, cards, etc.
- 209R (7133) Local tag with the same structure of the subfields in tag 009P/03 (4083), described in 'c. on-line monograph'.
This local counterpart of fields 4083 is used for off-line material when it is copied to a local document server, for deposit purposes and made available via the local library network through a 'local URL' tag.
- 248H (XXXX) (proposed) Local tag with the same information as in 048H (4251) but in encoded form.
This is technical information concerning installation and de-installation of the material, and possibly also for conversion purposes. This tag is specifically proposed for use by deposit material.

b) Off-line serial (ie CD-ROM journal)

- 011@ (1100)
- 016A (1101)
- 039Q (4262)
- 048H (4251) System requirements. For on-line: Internet, browser, viewer, telnet client, FTP client, WWW client, etc.
- 209R (7133)
- 248H (XXXX).

c) On-line monograph (ie dictionary on the Web)

- 011@ (1100)
- 016A (1101)
- 039Q (4262)
- 048H (4251) This field contains information on the original version of the on line publication (on the side of the publisher).
- 009P/03 (4083) \$S Coded information: available: free or by account; \$0 file format: HTML, pdf, ps, wp51, etc.; \$a URL; \$b file name ; \$c path; \$d number of files, bytes; \$e compression format ; \$m type of connection; \$n port number/protocol; \$o gopher type; \$p name of computer host; \$q IP address computer host; \$r location host; \$s mail address host; \$t mail address person.
This field contains location and file information. URL and file format are proscribed sub-fields. Most other subfields are not mandatory.
- This is a repeatable field and subfields are also repeatable to cater for multiple URLs and multiple formats of the same publication.
- It is possible that after an evaluation this tag will be totally revised.
- 009P/04 (4084) \$0 file format \$a the URL of an image; \$b file name; \$c path; \$d number of bytes; \$e compression format

Field 4084 contains location and file information of the images linked to the document.

The purpose of this field is to present an image which is linked to the actual document within the bibliographic description. This was a special feature PICA introduced for the WebDOC project.

N.B. If the on-line resource has an image format itself, the information of this image is entered in field 4083.

- 209R (7133)
- 209S (7134)

These local tags have the same structure as tags 4083 and 4084 and are used when the on-line resource which originally could be found elsewhere on the Internet is copied to a local document server.

- 248H (XXXX).

d) On-line serial (ie electronic journal on the Web)

- 011@ (1100)
- 016A (1101)
- 039Q (4262)
- 048H (4251)
- 009P/03 (4083)
- 009P/04 (4084)
- 209R (7133)
- 209R (7234)
- 248H (XXXX).

On-line articles

In the PICA system it is possible to describe on-line articles. For lending systems the following tags are important:

- 031A (4070) \$ text; \$d volume; \$j year; \$e issue; \$c month; \$b day; \$h pages; \$g total amount of pages.
- 039B (4241) PPN (Pica Production Number) with reference to the title of the journal.

e) On-line resource (i.e. an organisation's home pages)

With the tags mentioned above and the tags specified for the traditional publications it is possible to catalogue home pages in the PICA system. Home pages are described as on-line monographs or serials.

3. Which MARC (or equivalent) fields do you use to hold the data element?

- a) Off-line monograph
- b) Off-line serial (ie CD-ROM journal)
- c) On-line monograph (ie dictionary on the Web)
- d) On-line serial (ie electronic journal on the Web)
- e) On-line resource (ie. an organisation's home pages)

See answers to question 2.

4. What data would you like to include in the record that you cannot find, or have difficulty finding from the publication?

- a) Off-line monograph (ie CD-ROM encyclopaedia).
- b) Off-line serial (ie CD-ROM journal)
- c) On-line monograph (ie dictionary on the Web)
- d) On-line serial (ie electronic journal on the Web)
- e) On-line resource (ie. an organisation's home pages)

1- Importance of the source of the description

Traditionally the content of a publication is regarded as primary information source of the description. Bibliographic information embedded in the publication is considered more authoritative. We feel this is also true for electronic publications. If metadata is obtained through

other sources (for example through contact with the author) it is put between straight brackets [] in the ISBD presentation.

2- terms and condition/ access information: no special fields

At the moment we use several annotation fields. The format for electronic publications is still being developed. It is well tested and if necessary it is extensible.

For information that we cannot put in special fields, we use tag 4201. Often it concerns information that tends to become outdated quickly. Such information requires frequent updates, which we feel is not feasible in our current manual cataloguing practice. Updates should ideally be made by the producer or third party that offers the product/service.

- 4201 General note area.

General remarks related to the described document can be entered in the general note area.

Example:

4201 Database is available free of charge as part of the GPO Access System. Internet users should telnet to leocator.access.gpo.gov and login as leocator. Dial in users should use communications software and modem to call (202) 512-1661. No password is required; a screen message will then instruct the user in searching the database

4201 You can subscribe by sending a message through e-mail to: listserver@nix.surfnet.nl, the first line should read the command: SUBSCRIBE Periodical First name Surname

4201 Back issues of xxx are also available as follows; by e-mail: mailserv@nic.surfnet.nl. Send a message with the following contents: help index xxx

4201 If your institute has a licence agreement with SURFnet, you can order the SURFnet Guide 94/95 through your local system manager -Instellings Contact Persoon- (ICP). The Guide 94/95 is also available in the bookshop for f37,50 (including BTW). The ISBN is 90-74719-01-5.

3- technical information: encoded fields

At present cataloguing practice within the KB is to make a copy of the text on system requirements provided by the publisher/distributor into field 4251, which is an annotation field with free text.

There is however for the development of our electronic deposit (DNEP) an urge to record all relevant technical information concerning a particular electronic publication in a SYSTEMATIC way (encoded field proposal). In the processing model for electronic publications for DNEP, the following activities are aimed to ensure the publication will be readable in generations to come:

- installation
- description
- de-installation
- technical description (in encoded form preferably).

The provided system requirements by the publishers + the additional technical data necessary for installation and de-installation are to be recorded in such a way that in a future system, the codes can active software libraries and start applications automatically.

5. How have you resolved or attempted to resolve the difficulty?

By developing test-beds.

6. What library services will use metadata for electronic publications?

Examples: facilitating access (giving terms & conditions, access instructions); record supply.

1. OPAC (local retrieval system)
2. Union Catalogue (NCC) (national retrieval system + ILL)
3. Delivery of electronic resources e.g. published by universities (WebDOC)
4. List of electronic publications as a by-product of the national bibliography
5. (possibly in the future) alerting services.

7. What metadata elements do you consider will be required for electronic publications?

Examples: description of content (including mention of sub-units, contents pages); relation to other documents

- a) Off-line monograph (ie CD-ROM encyclopaedia)**
- b) Off-line serial (ie CD-ROM journal)**
- c) On-line monograph (ie dictionary on the Web)**
- d) On-line serial (ie electronic journal on the Web)**
- e) On-line resource (ie. an organisation's home pages)**

See answers to question 4 on 1) timely info and 2) technical info.

Also content information will be required - but this will not necessarily need to be different from the way printed material is disclosed by subject. But this issue might need looking into furthermore: what enrichment can national libraries provide here, that publishers don't? Controlled vocabularies and thesauri?

At the moment we are working on a conversion model to map the minimum requirements of metadata we would like to receive from publishers to the extended PICA format.

A working group of PICA participants from Germany and The Netherlands is composing a checklist of metadata (inspired by the 'Dublin Core'). The metadata of this list which could be embedded in a publication are to be automatically converted to the contents of tags in the PICA format. For the moment the proposal for the list is:

Labels	Structure	PICA+
Author	<Name>, <First name> <(Addition)>	\$d<First name>\$a<Na me\$y<Additio n> 028C/0X
Next author (repeat.)	<Name>, <First name> <(Addition)>	\$d<First name>\$a<Na me\$y<Additio n>
Corporation as author	Free text	029F/01 \$a
Next corporation	Free text	029F/02 \$a
Title	Free text	021A \$a
Place of publication	Free text	033A \$p
Publisher or host	Free text	033A \$n
Next place of publication	Free text	033B/01 \$p
Next publisher or host	Free text	033B/01 \$n
Date, year	dd-mm-yyyy	011@ \$ayyyy\$ndd_ mm[converted from table]_yyyy
Language	In English: <free>	010@ \$axx [converted from table]
Next language (rep.)	In English: <free>	010@ \$bxx [converted from table]
Edition/update	Free text	032@ @a
Form	Multiple choice	009P/03 \$0
URL	Structured	009P/03 \$a
Subject keywords	In English: <free>	144Z \$a
Next subject keywords	In English: <free>	144Z \$a
Abstract	Free text in English	047I \$a

For clearness' sake: this a proposal of KB and it is not yet (today is 15 October 1996) discussed in the working group.

Author

Next author

A slight structuring is necessary. The inversion of surname and first indicated by, (comma) is I suppose in common use. With this interpunction we can convert in the majority of cases to the PICA format. This is mandatory for adequate indexing of modern western author names.

E.g.:

Author:Bohemen,Francien van;converted to:\$d Francien van\$aBohemen(3XXX Francien van
@Bohemen); indexed as: Bohemen, Francien van

Author: Van Bohemen, Francien; converted to: \$dFrancien\$aVan Bohemen (3XXX Francien van
@Bohemen); indexed as: Van Bohemen, Francien, and as: Bohemen, Francien van

There must be an opportunity to add different kind of information: on the author as person and about the relation of the author to the publication: composer, editor, translator, etc. I trust it is possible to

educate the editors of WebCat publications to enter information between (), parentheses. Conversion of this additional information to subfields f, h or i is possible for different kinds of displaying. I propose the scarcely used \$y (addition of title), in cataloguing format to precede by _+_. If there are objections to this solution a new subfield is needed.

E.g.:

Author: Bohemen, Francien van (projectmanager WebDoc); converted to: \$Dfrancien van \$aBohemen\$yprojectmanager WebDoc(3XXXFrancienvan@Bohemen +projectmanager WebDoc). It is the responsibility of the cataloguer to bring the information of the addition in the convenient place of the bibliographical record. Indication as editor, etc. are to be brought in the statement of responsibility of the title field.

The author label is repeatable. Next authors are converted to the field 028C/0X (310X (N), 301X (D).

Corporation as author

It is not convenient to prescribe the WebDoc editors to create any kind of structure in names of corporations. The contents of 029F/01 \$a is word based indexed. Therefore conversion from a free text of this label to tag 029F/01 is acceptable.

E.g.:

Corporation: Department of New Testament Studies of the Faculty of Theology; Leiden University; converted to: 029F/01 \$department of New Testament Studies of the Faculty of Theology; Leiden University (3121 Department ...); indexed as: department, new, testament, studies, faculty, theology, Leiden, university. After conversion of the contents of the label correct indexing is assured for the time being. In a later stage a cataloguer can transform these date according the bibliographical rules.

The corporation (as author) label is repeatable.

Title

The major problem of the conversion of the label 'Title' to tag 021A \$a is the indication of Non Sorting Beginning and Non Sorting Ending. My premise is that we can not ask WebDoc editors to determine articles as non sorting. Accepting this the question is: does PICA have available, or is it possible to develop software to exclude the opening articles of a title from sorting? Is it acceptable to have for a (short?) time titles incorrect indexed? E.g.: De @~civitate Dei, instead of: ~De civitate Dei; The ou cafe, instead of: The ou cafe? Or do we want more intelligent software with a connection to the language? But the indication of the languages is uncertain.

Because of the relative low amount of titles with a wrong NSB/NSE indication I suppose to make simple software: with only a simple checklist of the articles of the most important languages in the WebDoc area (English, German, Spanish, Italian, Dutch) a large majority of titles will have the right NSB/NSE indication.

Place of publication**Publisher or host**

I propose to define two labels for publisher and place of the publisher and to make the couple of the two labels repeatable. To distinguish hosts from publishers, host are to be indicated by: [host].

E.g.:

- Place of publication: Leiden Publisher or host: PICA
- Next place of publication: Utrecht
- Next publisher or host: SURF [host]; converted to:
- 033A \$pLeiden\$nPICA (4030 Leiden: PICA)
- 033B/01 \$pUtrecht\$nSURF [host] (4031 Utrecht: SURF [host]).

Date, year

In the PICA databases the indexing and **display of year (or more specific date)** is complex. When the WebDoc editors are following the simple rule 2 numerals for the day-2 for the month-4 for the year we realise a convenient conversion to PICA+. In this case it will be necessary to make a concordance of the numbers of the month and the abbreviation prescribed in the PICA Guidelines for Cataloguing On-line Resources: 01 =Jan.; 02 =Feb.; 03 =Mar.; 04 =Apr.; 05 =May; 06 =June; 07 =July; 08 =Aug.; 09 =Sept.; 10 =Oct.; 11 =Nov.; 12 =Dec.

E.g.:

Date, year: 29-09-1996; converted to: 011~at \$al996\$nO1 Sept. 1996 (1100 1996 \$ 01 Sept. 1996).

Language

Next language

The PICA language codes are requested for the system. For practical reasons it is impossible to ask WebDoc editors to use the PICA language code list. Therefore a concordance English/German terms to the language code is required. For the English retrieval system of the Dutch Bibliography a machine readable concordance is made: aco-Acoli-Acoli ...zwe-Zweeds-Swedish. This concordance, or a part of it with for our purpose the most important languages, can be used with addition of variant terms: old Greek= grk; Classical Greek=grk; Modern Greek=grn; New Greek=grn. When the label language contains a term not existent in the concordance the PICA language code onb (onbepaald, undefined) is to be generated. [N.B.:the examples of codes are from the Dutch situation.]

E.g.:

- Language: Latin
- Next language: Middle High German Next language: English;
- converted to:
- 010@ \$alat\$bdmh\$beng
- Language: Berlin dialect; converted to: 010@ \$aonb

Edition/update

Edition/update is here in the traditional meaning: as in paper publications: edition, printing, Ausgabe, Auflage, editie, druk, etc. In electronic publications it is possible that the indication of edition/update is equal to the term of the date of publishing. It is also possible that there are connections with the next label Form. E.g. the contents of the second edition in the Postscript form can be the same of the third edition of the HTML form. The cataloguer must combine the data from the labels Year, Edition and Form for creating bibliographical data according the rules.

Form

A kind of standardisation of the terms used is desired. I refer to the description of the elements in the Dublin Core: 'examples might include Postscript-II document, Windows 3.1 executable file, HTML file, or WordPerfect 6.1 document. The PICA subfield \$0 from tag 009P/03 precede subfield \$a, the URL in the Dutch set of subfields in tag 009P/03.

E.g.: 009P/03 \$0html\$ahhttp://www.oclc.. (4083 html =A http://www.oclc..)

URL

The URL is constructed according delimited rules. The cataloguer has to control the data converted from the label to the PICA field.

Subject keywords

Next subject keywords

For the moment I suppose to convert the contents of the label Subject keywords to a tag form the local level . The local cataloguer judges the proposal of WebDoc editor, if possible in relation with the abstract, and brings the subject indexing to the common level (basis classification, 045Q). The German participants are inspecting the possibilities of aliasing free subject terms to authorised term.

Abstracts

The maximum length of field 0471 \$a must be the same or greater than the length specified for the contents of the label.

8. Will the records for electronic publications need to be integrated in your existing systems?

Yes, the records of electronic publications produced in the PICA format are in the same database with the other title records. It must be possible to have both kind of records in the OPAC's of the library and e.g. in files of the National Bibliography. A total integration is required.

9. Will the metadata need to be manipulated by particular protocols e.g. Z39.50, DE, etc.?

PICA exploits since the end of 1995 an Z39.50 entry to the central database GGC. In view of this it is important to investigate the pros and cons of manipulating metadata by the Z39.50 protocol in comparison with the use of the PICA protocol.

Nowadays we buy tapes with bibliographic records which are integrated in the common database maintained by Pica. As output, tapes are sent to our CD-ROM production company in the United States. Possibilities to use FTP for transfer of the records of the national bibliography for the CD-ROM are investigated.

10. Please add any further comments about your experiences in cataloguing electronic publications.

The cataloguing of CD-ROM and diskettes does not present specific problems. Except for the difficulty in some cases to judge whether a hybrid publication (book + diskette) is mainly printed or mainly an electronic publication.

Concerning on-line resources, especially homepages: At this moment three cataloguers are making bibliographic descriptions for home pages. One for the scientific collection, one for the union catalogue and one for the national bibliography. Further, since a common cataloguing systems is used they can also extract records from the common database and they can see how cataloguers in other organisations, especially one university library catalogue e.g. homepages. The problems with homepages are : author unknown, publisher unknown, date of publication unknown, change of even disappearance of the URL.

BOff-line electronic publications are catalogued and published in the national bibliography.

11. Wish list - anything else?

Not so much wishes but some additional thoughts that have come up here when tackling the cataloguing issue:

1. Importance of the presentation

For the National Bibliography and other metadata products and services of the library, the consistent and user-friendly presentation of records on screen or on paper is very important. ISBD plays therefore an important role. ISBD is a standard for presentation of bibliographical records. In the international library community a proposal for an ISBD(CF), the presentation of metadata of computer files, has been commented upon. In December 1996 the final version is expected.

The terminology of CF (computer file) has already been adapted to ER (electronic resources). In the meantime we have defined ourselves an 'ISBD'-like presentation for 'On-line Resources' and PICA has implemented this in its system. When the new ISBD(ER) will become a standard PICA and KB will as a matter of course adhere to this new standard.

2. Linking between different versions of electronic publication

We would wish to link all versions of the same publication transparently and present the user with all versions available. Another reason for linking different versions is the possible economy of cataloguing efforts. This is extremely difficult however. First of all one must be sure 2 versions of the same publication are exactly the same content wise!

At the moment we have a linking facility between records of the printed and electronic versions of a publication. The record of the printed publication is the master record. The record of the electronic version is appended via a PPN-link. This is the internal PICA record number (PPN). This linkage does not work properly due to indexing complications.

Links between records of different versions of an electronic publication (CD-ROM/on-line) is not possible. Different formats of the same electronic publications are catalogued in the same record: the URL-tag is repeatable and caters for multiple locations and formats of the same publication. Links between journal and articles of the journal (whole/part relationship) should also be realised in the system. In the PICA system this is also done with the PPN.

B.3 Norway response**1. Please indicate the cataloguing system you use (for example, UNIMARC), and which cataloguing rules you apply.**

Cataloguing system: BIBSYS-MARC, based on the USMARC

Cataloguing rules: Norwegian cataloguing rules, based on AACR II:

Katalogiseringsregler : Anglo-American cataloguing rules, second edition / oversatt og bearbeidet for norske forhold ved Inger Cathrine Spangen. - Oslo : Norsk bibliotekforening, 1983. - 610 s. - ISBN: 82-990932-0-1

In addition a Norwegian supplement for electronic document (machine readable files), which is mainly used on off-line documents (not totally satisfactory for on-line documents):

Maskinlesbare filer : reviderte regler for beskrivelse med tilhørende eksemplarsamling ; samt Revisjoner og tilføyelser til Katalogiseringsregler 1983 / ved Inger Cathrine Spangen. - Oslo : Norsk bibliotekforening, 1990. - 86 s. : ill. - ISBN: 82-90790-01-5

And we are using:

*Cataloguing internet resources. A Manual and Practical Guide. Nancy B. Olson, editor.(OCLC 1995
ISBN 1-55653-189-3. - <http://www.oclc.org/oclc/man/9256cat/toc.htm>*

2. What data elements do you record for an electronic publication in addition to those you record for a traditional publication? Also, what data elements do you not record that you would use for a traditional publication? (Sample records may be attached.)

[I have listed all the data elements we record for electronic documents. Data elements in **bold-faced type** in addition to those we record for traditional publications is written in **bold-faced type**]

a) Off-line monograph (i.e. CD-ROM encyclopaedia)

- **File characteristics**
- **System requirements**
- General material designation
- Coded information for type of document
- Physical description.

b) Off-line serial (i.e. CD-ROM journal)

Same as a) (As far as we know at the moment, we have not catalogued any of these so far)

c) On-line monograph (i.e. dictionary on the Web)

- System requirements
- File characteristics
- Mode of access (e.g. Internet)
- Location. Information on the original URL (the publishers URL). For legal deposit documents, we might want to add the local URL of the copy stored in the deposit library. (We have so far no actual practice in cataloguing deposited on-line material, but we intend to start deposit of electronic journals this autumn and our plans is to register both the publisher URL and the local URL)
- General material designation
- Coded information for type of document.

We do not add any links to printed version (if any) of the documents.

d) On-line serial (i.e. electronic journal on the Web)

Same as c) and...

- Unique title, additional information.

e) On-line resource (i.e. organisation's home pages)

Same as c) (As far as we know at the moment, we have not catalogued any of these so far)

3. Which MARC (or equivalent) fields do you use to hold the data element?

[I have listed all the data elements we record for electronic documents. Data elements in **bold-faced type** those we record for traditional publications is written in **bold-faced type**]

a) Off-line monograph (i.e. CD-ROM encyclopaedia)

- ***256\$a** (File characteristics)
- ***500\$a** (System requirements)
- ***245\$h** (General material designation)
- ***008\$a** (Coded information for type of document)
- ***300\$a\$b** (Physical description).

b) Off-line serial (i.e. CD-ROM journal)

Same as a) (As far as we know at the moment, we have not catalogued any of these so far)

c) On-line monograph (i.e. dictionary on the Web)

- *256\$a (File characteristics)
- *500\$a (System requirements)
- *500\$a (Mode of access)
- *856\$u\$z (Location)
- *245\$h (General material designation)
- *008\$a (Coded information for type of document).

We do not add any links to printed version (if any) of the documents.

d) On-line serial (ie electronic journal on the Web)

Same as c) and...

- *130\$q (Unique title, additional information).

e) On-line resource (i.e. an organisation's home pages)

Same as c) (As far as we know at the moment, we have not catalogued any of these so far)

4. What data would you like to include in the record that you cannot find, or have difficulty finding from the publication?**a) Off-line monograph (ie CD-ROM encyclopaedia)**

- System requirements and other technical information are sometimes difficult to find and recording the data would be easier if they were listed «somewhere» in a standard format.
- Information on access conditions.

b) Off-line serial (i.e. CD-ROM journal)

Same as a)

c) On-line monograph (i.e. dictionary on the Web)

- The «size» of the document. What (and how many) files/records does it consist of.
- A listing of links to other documents.

d) On-line serial (i.e. electronic journal on the Web)

Same as c)

e) On-line resource (i.e. organisation's home pages)

Same as c)

5. How have you resolved or attempted to resolve the difficulty?

By asking the publishers and/or the staff in the Information technology department and/or librarians cataloguing the same kind of material in other libraries.

6. What library services will use metadata for electronic publications? Examples: facilitating access (giving terms & conditions, access instructions); record supply.

All library catalogues, the metadata will probably have to be refined to meet the requirements of the National bibliography.

7. What metadata elements do you consider will be required for electronic publications? Examples: description of content (including mention of sub-units, contents pages); relation to other documents**a) Off-line monograph (i.e. CD-ROM encyclopaedia)**

In addition to data elements recorded for traditional documents and elements listed in «2» and «4»:

- Description of content, e.g. subject keywords.
- Version.

b) Off-line serial (i.e. CD-ROM journal)

Same as a)

c) On-line monograph (i.e. dictionary on the Web)

Same as a) and:

- Listing of relation (links) to other documents and links within the document.

d) On-line serial (i.e. electronic journal on the Web)

Same as c)

e) On-line resource (i.e. an organisation's home pages)

Same as c)

8. Will the records for electronic publications need to be integrated in your existing systems?

Yes.

9. Will the metadata need to be manipulated by particular protocols e.g. Z39.50, EDI, etc.?

Yes, certainly by Z39.50. And by other relevant protocols (even if we not use e.g. the EDI protocol).

10. Please add any further comments about your experiences in cataloguing electronic publications.

11. Wish list - anything else?

The quality of the metadata elements should be ensured, especially by controlled vocabulary, e.g. name of persons and institutions, subject keywords.

B.4 Biblioteca Nacional**1. Please indicate the cataloguing system you use (for example, UNIMARC), and which cataloguing rules you apply.**

For the time being, the Biblioteca Nacional only catalogues off-line publications, both monographs and serials. Deposit law requires deposit of off-line publications but not of on-line publications.

Monographs are catalogued in the Computer File module of our cataloguing system, but not serials, due to problems in the linkings between the computer files and the paper records. Serials can be catalogued by the Serials Cataloguing Section provided that they are received through legal Deposit, or by the Spanish ISSN Centre if the publisher asks this Centre for the assignment of ISSN. The same serials format is used for cataloguing for the Biblioteca Nacional database (ARIADNA). Although the fields used in MARC and ISSN sometimes differ, a conversion program from ISSN to MARC has been developed at the Biblioteca Nacional.

The cataloguing of on-line publications has not started yet. If an editor asks for an ISSN for a given on-line journal, he is required to record the information in a diskette, which is used as the cataloguing source.

The cataloguing system we use is IBERMARC, a variant of the MARC format. The cataloguing rules we apply are the Spanish Cataloguing Rules, based on the ISBD.

2. What data elements do you record for an electronic publication in addition to those you record for a traditional publication? Also, what data elements do you not record that you would use for a traditional publication? (Sample records may be attached.)

2. a) Off-line monograph: Specific material designation, File characteristics, Technical data, File type, Related files, Other formats available, Linking entry complexity note. (Sample record attached)
- b) Off-line serials: Specific material designation, System requirements, Linkings.
- c) d) e) They have not been catalogued so far

3. Which MARC (or equivalent) fields do you use to hold the data element?

3. a) Off-line monograph: 245\$h, 256, 538, 516, 530, 580.
- b) Off-line serial: 245\$h, 500, 775

4. What data would you like to include in the record that you cannot find, or have difficulty finding from the publication?

4. In the case of monographs, sometimes is difficult to determine the technical data (753), the ISBN and the Legal Deposit numbers because they are not shown in the disc label or elsewhere.

In the case of serials we have problems to know if the issue we have in hand is the first one, because it is produced not by an established publishing house but by some kind of firm that has experience in data processing and not in publishing.

In addition, questions as "statements of responsibility" or "edition" are sometimes difficult to determine.

5. How have you resolved or attempted to resolve the difficulty?

In most cases we make a phone call to the publisher and ask for information about frequency, date of publication, title, etc.

6. What library services will use metadata for electronic publications? Examples: facilitating access (giving terms & conditions, access instructions); record supply.

Most library services will use metadata. A problem to be regarded with deposited electronic publications is copyright and a possible unfair use.

7. What metadata elements do you consider will be required for electronic publications?

Examples: description of content (including mention of sub-units, contents pages); relation to other documents

- Access data (location, format, etc.)
- Data about edition or version, some kind of "control number"
- Description of content
- Frequency or date of issue of next expected issue
- Identification of the "Publisher"

- Preservation data (for preservation management purposes)
- Terms and conditions (restrictions, licence information)
- Subject terms, classification.

8. Will the records for electronic publications need to be integrated in your existing systems?

Yes. The National Library wants to have a single database.

9. Will the metadata need to be manipulated by particular protocols e.g. Z39.50, EDI, etc.?

Yes. At least, Z39.50, ISO 10162/10163, EDI, and any other protocols established by the DG.XIII.

10. Please add any further comments about your experiences in cataloguing electronic publications.

We have few experience on this type of publications but we are very interesting in all problems with this material.

B.5 The British Library

The responses to the questionnaire have been gathered from three different sources. They have been collated into this document.

1. **Cataloguing system and tools used** - Washington Library Network system (to be replaced in the near future, UKMARC, AACR2r).
2. **Fields that differ between electronic and print records** - For all types of electronic resource a notes field giving system requirements is needed (**542**). Title notes (**514**) are also needed to give the source of the title.

An Internet resource should have an **856** field giving the URL. Currently, for remote resources, either serial or monograph, no physical description field is given (**300**).

Electronic resources should also have a **258** field describing file characteristics which is not currently featured in the UKMARC manual.

3. **MARC fields for electronic publications:** - If the data element is taken as being the information in the record which indicates that the item is an electronic resource, where this appears will vary from item to item.

For physical format fields, the item is described in the **300** field, as follows: *300.00 \$f1\$ncomputer laser optical disc* to indicate 1 CD-ROM.

Absence of a **300** field and the presence of an **856** field or notes fields (**5XX**) indicating a service provider and login name and password, or system requirements will indicate that the item is not locally held.

The **008** field indicates in both cases that the item is an electronic resource, as the **008 \$p** contains a **K**.

Generally speaking, home pages are treated in the same way as on-line monographs are treated, and a date is applied which reflects either the time of cataloguing, or if given, the most recent date of modification.

The General Material Designation [computer file] is not applied as part of library practice at the moment, but may be in the future. If it were to be applied, it would be present in all the types of records mentioned.

4. **Data which is difficult to find** often includes the following (true for all types of resources, but in particular the on-line variety!) -

Publisher name and location and clear statements of responsibility which describe what an individual's contribution has been.

On some resources (both off-line and on-line) it isn't easy to determine the correct title, as variants on boxes, accompanying information and internal sources often conflict. Some items have no internal title at all (not usually true for CD-ROMs, but certainly true for less "public friendly" productions like databases).

Version information is not always clear, particularly when an updated software version may be released with the manual for the previous version.

Sometimes the stated systems requirements differ between internal and external sources, often not distinguishing between required and preferred hardware and software.

5. **Resolving the difficulties.** In general, cataloguing rules do not allow you to assume anything, so if information is not available, statements of responsibility are left blank and the appropriate terms for "not known" are inserted.

If a title is not present, the cataloguer is allowed to create one that they feel is descriptive of the item and a note is added to say that it is a given title.

A trial project cataloguing off-line publications has been run and a larger one is planned in support of the extension of legal deposit to non-print items. Provisional policy guidelines are being developed.

6. **Library services:** - Obviously as the main cataloguing department, we need to be providing metadata for the collections, if the collections are collecting electronic items. Whether we use metadata that

has already been encoded as our base record, or whether we use metadata that is made available to enhance our own metadata is a policy issue which as yet hasn't been dealt with.

In addition to facilitating access and record supply, these library functions will also be served by the appropriate metadata -

- selection
- acquisition
- preservation and archiving.

7. Metadata elements that will be required:- File size and type of content is always useful information (whether there are audio/graphics files, what type of files these are, whether the product is designed to interactive or not).

On-line resources are always more reliably catalogued if they include a date/time stating when the last update occurred.

For both types of serial, it would be useful to know what the frequency is, and whether this is regular, or whether new papers are posted as soon as they are approved for publishing (more likely to be the case with on-line journals).

It would also be helpful to know whether an item bears any relation to another item as this sort of information is included in records when it is available.

For works which already exist in a different form, (e.g. the digitisation of a novel) it would be helpful to know the extent and type of input from individuals other than the author(s) of the original work, to identify the additional access points.

Off-line monograph: contents page data; revision/updating arrangements (e.g. via the Web); multimedia content; hardware/software requirements; networkability

Off-line serial: contents page data; abstracts; relationships with other publications; hardware/software requirements; networkability; frequency and regularity

On-line monograph: contents page data; revision/updating arrangements (e.g. via the Web); multimedia content

On-line serial: contents page data; abstracts; relationships with other publications; frequency and regularity; policy regarding timing of appearance of new papers (at regular intervals or as soon as paper approved).

On-line resource: frequency of updating

For the British National Bibliography a summary of the contents would be useful.

8. Integration into existing systems: -Yes, as far as practicable and to the same basic bibliographic and format standards. The library is aiming for more uniform access to its collections, so any additional records would be expected to integrate into the library-wide system.

9. Manipulation by other protocols:- They may need to be manipulated by Z39.50 if the library is collaborating on Z30.50 projects. Also probably MARC, HTML

10. Comments and Wish list:- The most important issue for us at the moment is the whole problem of installing an item to catalogue it and not having clearly defined instructions to de-install it. Simply deleting the files is not sufficient, as there are sometimes attachments which disable the CD-ROM or floppy disk for a second installation, and sometimes floating files are left clogging up disk space on the PC. This makes their use in the Reading Rooms impossible. Of 29 items included in the investigation, only 1 had clear instructions in the user handbooks that often accompany the item. Much time can be wasted in looking for the relevant instructions. Quick installation procedures are desirable - the faster the installation process, the faster we can catalogue these types of items, and certainly at the moment, it takes longer to catalogue an electronic item than it does to catalogue a book.

Attachments include 3 examples of serial records, all of which are non-BNB because the current exclusion policy for BNB states that items which are predominantly computer files should not be included. Also attached are 3 monographic records and 2 for kits, also non-BNB for the same reason. (BNB records would have LCSH subject headings, Dewey Classification and availability information.)

ANNEX C. UKMARC FIELDS FOR CATALOGUING ELECTRONIC PUBLICATIONS

Already in manual

542 Mode of use note (machine-readable data files) (rule 9.7B15)

Examples

1. 542.00 \$aMode of use: On-line video or teletype terminal or with a small dedicated computer
(e.g. PDP8)

Approved, to be added to manual in an Update

856 Electronic Location and Access

The 856 field contains the information required to locate an electronic item. The information identifies the electronic location containing the item or from which it is available. It also contains information to retrieve the item by the access method identified in the first indicator position. The information contained in this field is sufficient to allow for the electronic transfer of a file, subscription to an electronic journal, or logon to an electronic resource. In some cases, only unique data elements are recorded which allow the user to access a locator table on a remote host containing the remaining information needed to access the item.

INDICATORS DEFINITION

1st	2nd	
0	0	Email
1	0	FTP
2	0	Remote login (Telnet)
3	0	Dial-up
7	0	Method specified in subfield \$y

SUBFIELD CODES

\$a	Host name
\$b	Access number
\$c	Compression information
\$d	Path
\$f	Electronic name
\$g	Electronic name--End of range
\$h	Processor of request (NR)
\$i	Instruction
\$j	Bits per second (NR)
\$k	Password (NR)
\$l	Logon/login (NR)
\$m	Contact for access assistance
\$n	Name of location of host in subfield \$a (NR)
\$o	Operating system (NR)
\$p	Port (NR)
\$q	File transfer mode (NR)
\$r	Settings (NR)
\$s	File size
\$t	Terminal emulation
\$u	Uniform Resource Locator
\$v	Hours access method available
\$w	Record control number
\$x	Non-public note
\$y	Access method (NR)
\$z	Public note

Notes

1. The field is repeated when the location data elements vary (subfields \$a, \$b, \$d) and when more than one access method may be used. It is also repeated whenever the electronic filename varies (subfield \$f), except when a single intellectual item is divided into different parts for on-line storage or retrieval.
2. Subfield \$b contains the access number associated with a host. It can contain the Internet Protocol (IP) numeric address if the item is an Internet resource, or a telephone number if dial-up access is provided through a telephone line. This data may change frequently and may be generated by the system, rather than statically stored. Subfield \$b may be repeated if all the other information in the field applies. A telephone number is recorded as follows: [country code]-[area code]-[telephone number]. Example: 61-49-215833 (a number in New South Wales, Australia); 1-202-7076237 (a number in the U.S., Washington, D.C.). If an extension is applicable, include it after the telephone number preceded by "x". Example: 1-703-3589800x515 (telephone number with extension).
3. Subfield \$f contains the electronic name of a file as it exists in the directory/subdirectory indicated in subfield \$d on the host identified in subfield \$a. Subfield \$f may be repeated if a single logical file has been divided into parts and stored under different names. In this case, the separate parts should constitute a single bibliographic item. In all other cases, a file that may be retrieved under different filenames contains multiple occurrences of field 856, each with its corresponding electronic name in subfield \$f. A filename may include wildcard characters (e.g., "*" or "?") if applicable, with a note in subfield \$z explaining how files are named. NOTE: Filenames may be case sensitive for some systems. This subfield may also contain the name of the electronic publication or conference.
4. Subfield \$g contains the name of the last physical file in an ordered sequence when the resource is contained in more than one physical file. This subfield is used when repetition of subfield \$f is not possible.
5. Subfield \$h contains the username, or processor of the request; generally the data which precedes the at sign ("@") in the host address.
6. Subfield \$i contains an instruction or command needed for the remote host to process a request.
7. Subfield \$j contains the lowest and highest number of bits (binary units) of data that can be transmitted per second when connected to a host. The syntax for recording the number of bits per second (BPS) should be: -. If only lowest given: - ; If only highest given: -.
8. Subfield \$k is used to record general-use passwords, and should not contain passwords requiring security.
9. Subfield \$l is used to record general-use logon/login strings which do not require special security.
10. For informational purposes, the operating system used by the host specified in subfield \$a is indicated in subfield \$o. Conventions for the path and filenames may be dependent on the operating system of the host. For the operating system of the resource itself (i.e., the item represented by the title recorded in field 245), rather than the operating system of the host making it available, field 753 (Technical Details Access to Computer File³) subfield \$c (Operating system) is used.
11. Subfield \$p contains the portion of the address that identifies a process or service in the host.
12. Subfield \$q contains an identification of the file transfer mode, which determines how data are transferred through a network. Usually, a text file can be transferred as character data which generally restricts the text to characters in the ASCII (American National Standard Code for

³Soon to be proposed for UKMARC

Information Interchange) character set (i.e., the basic Latin alphabet, digits 0-9, a few special characters, and most punctuation marks). Text files with characters outside of the ASCII set, or non-textual data (e.g., computer programs, image data) must be transferred using another file transfer mode, usually binary mode

13. Subfield \$r contains the settings used for transferring data. Included in settings are: 1) Number Data Bits (the number of bits per character); 2) Number Stop Bits (the number of bits to signal the end of a byte); and 3) Parity (the parity checking technique used). The syntax of these elements is: -- If only the parity is given, the other elements of settings and their related hyphens are omitted (i.e., ""). If one of the other two elements is given, the hyphen for the missing element is recorded in its proper position (i.e., "--" or "--") The values for parity are: O (Odd), E (Even), N (None), S (Space), and M (Mark).
14. Subfield \$s contains the size of the file as stored under the filename indicated in subfield \$f. It is generally expressed in terms of 8-bit bytes (octets). It may be repeated in cases where the filename is repeated and directly follows the subfield \$f to which it applies. This information is not given for journals, since field 856 relates to the entire title, not to particular issues.
15. Subfield \$u contains the Uniform Resource Locator (URL), which provides electronic access data in a standard syntax. This data can be used for automated access to an electronic item using one of the Internet protocols. Field 856 is structured to allow the creation of a URL from the concatenation of other separate field 856 subfields. Subfield \$u may be used instead of those separate subfields or in addition to them. Subfield \$u may be repeated if the other information in the field applies.
16. If the record is for a system or service, the hours of availability are recorded in field 307 (Hours, etc.)⁴ and not subfield \$v.
17. Subfield \$y contains the access method when the first indicator position contains value 7 (Method specified in subfield \$y). This subfield may include access methods other than the three main TCP/IP protocols specified in the first indicator. The data in this subfield corresponds with the access schemes specified in Uniform Resource Locators (URL) (RFC 1738), a product of the Uniform Resource Identifiers Working Group of the IETF. The Internet Assigned Numbers Authority (IANA) maintains a registry of URL schemes and defines the syntax and use of new schemes; the British Library will include an authoritative list based on that standard in *USMARC Code List for Relators, Sources, Description Conventions*.

Examples

1. 856 10\$awuarchive.wustl.edu\$cdecompress with PKUNZIP.exe
\$d/mirrors2/win3/games\$fatmoids.zip\$xcannot verify because of transfer difficulty
2. 856 10\$uftp://path.net/pub/docs/urn2urc.ps
856 70\$uhttp://lcweb.loc.gov/catdir/semdigdocs/seminar.html\$yhttp
3. 856 20\$apac.carl.org\$b192.54.81.128\$mCARL Situation Room\$mhelp@CARL.org\$nCARL Systems Inc., Denver, CO\$v24 hours
4. 856 30\$alocis.loc.gov\$b140.147.254.3\$mlconline@loc.gov\$t3270\$tline mode (e.g., vt100)\$vM-F 6:00 a.m.-21:30 p.m. USA EST, Sat. 8:30-17:00 USA EST, Sun. 13:00-17:00 USA EST
5. 856 70\$uhttp://lcweb.loc.gov/catdir/toc/93-3471.html\$yhttp
6. 856 70\$uhttp://lcweb.loc.gov/catdir/semdigdocs/seminar.html\$yhttp
7. 856 20\$amaine.maine.edu\$nUniversity of Maine\$t3270

⁴ Soon to be proposed for UKMARC

8. 856 10\$awuarchive.wustl.edu\$dmirrors/info-mac/util\$fcolor-system-icons.hqx\$s16874 bytes
856 00\$akeptvm.bitnet\$facadlist file1\$s34,989 bytes\$facadlist file2\$s32,876 bytes\$facadlist
file3\$s23987 bytes
9. 856 20\$agopac.berkeley.edu\$mRoy Tennant
10. 856 30\$b1-202-7072316\$j2400-9600\$nLibrary of Congress, Washington, DC\$oUNIX\$rE-7-1
\$tvt100\$zRequires logon and password
11. 856 10\$aarchive.cis.ohio-state.edu\$dpub/comp.sources.Unix/volume
10\$fcomobj.lisp.10.Z\$qbinary
12. 856 10\$aunmvm.bitnet\$lanonymous
13. 856 10\$aseq1.loc.gov\$d/pub/soviet.archive\$fk1famine.bkg\$nLibrary of Congress, Washington,
D.C.\$oUNIX
14. 856 20\$amadlab.sprl.umich.edu\$nUniversity of Michigan Weather Underground\$p3000
15. 856 20\$apucc.princeton.edu\$nPrinceton University, Princeton, N.J.
16. 856 00\$auccvma.bitnet\$fIR-L\$hListserv\$isubscribe
17. 856 00\$b1-202-7072316\$j2400-9600\$nLibrary of Congress, Washington, DC\$oUNIX\$rE-7-1
18. 856 00\$b1-202-7072316\$j2400-9600\$nLibrary of Congress, Washington, DC\$oUNIX\$rE-7-1
19. 856 10\$asunx.loc.gov\$dLCPP04A\$f4A49751\$g4A49755
20. 856 10\$aharvarda.harvard.edu\$kguest
21. 856 00\$auicvm.bitnet\$fAN2
22. 856 10\$awuarchive.wustl.edu\$dmirrors/info-mac/util\$fcolor-system-icons.hqx
23. 856 00\$akentvm.bitnet\$facadlist file1\$facadlist file2\$facadlist file3
24. 856 20\$aanthrax.micro.umn.edu\$b128.101.95.23
- Host name and Internet Protocol numeric address*
25. 856 30\$b1-202-7072316\$j2400/9600\$nLibrary of Congress, Washington, DC\$oUNIX\$rE-7-1
\$tvt100\$zRequires logon and password
- Dial-up numbers with related settings for terminal emulation*
26. 856 10\$amaine.maine.edu\$cMust be decompressed with PKUNZIP\$fresource.zip
27. 856 10\$awuarchive.wustl.edu\$d/aii/admin/CAT.games\$fmac-qubic.22.hqx

In the December 96 consultative document

258 Computer File Characteristics (NR)

This field is used to record characteristics pertaining to a computer file. It may contain information about the type of file (e.g., Computer programs), the number of records, statements, etc. (e.g., 1250 records, 5076 bytes).

Indicators Definition AACR2

1st 2nd

0

Subfield Code

\$a Computer file characteristics (NR) 9.3

Examples

- Example 11**

 1. 258 00\$aComputer data (2 files: 876,000, 775,000 records).
 2. 258 00\$aComputer programs (2 files: 4300, 1250 bytes).
 3. 258 00\$aComputer data (2 files: 800, 1250 records) and programs (3 files: 7260, 3490, 5076 bytes).
 4. 258 00\$aData (1 file: 350 records).

307 Hours, etc. (NR)

This field contains chronological information identifying the days and/or times an item is available or accessible. It is used primarily in records for electronic resources.

Indicators Definition

1st 2nd

0	0	No information provided
0	8	No display constant generated

Subfield Codes

\$a Hours (NR)

\$b Additional information (NR)

Notes

1. Subfield \$a contains information that identifies the days and/or hours an item is available or accessible. Informal references to the A.M. and P.M. time references as well as time zone can be given in this subfield if required.
2. When displayed or printed as a note, hours, etc information is in some instances preceded by an introductory term or phrase that is generated based on the first indicator value.

Examples

1. 307 00\$aM-F, 9AM-10PM
2. 307 00\$aTu-F, 10-6; Sa, 1-5, USA PST
3. 00\$aM, 8:30-6:00, Tu, 8:30-7:00; W-F, 8:30-6:00;\$bnot available on weekends.
4. 80\$a8:00 p.m., Tu-F; 5:00 and 9:00 p.m., Sa; 2:00 and 7:00 p.m., Su (all times, EST)
5. 00\$aM-F, 6:30am-9:00pm (EST); \$bwith brief interruptions for periodic update/backup of data
6. 00\$aDaily, 7am-7pm;\$text files only
7. 00\$aM-F, 6:30 AM to 9:30 PM, Sa, 8:00 AM to 5:00 PM, Su, 1:00 PM to 5:00 PM; \$bclosed on national holidays (all times are EST or ESDT)

753 System Details Access to Computer Files

This field contains information relating to the type of machine, operating system, and/or programming language used with computer files and accompanying material. This kind of added entry is assigned to give access to the bibliographic record which otherwise would not be possible and to facilitate the capability of selecting and arranging records for production of printed indexes.

Indicators Definition

1st 2nd

0 0

Subfield Codes

- \$a Make and model of machine (NR)
- \$b Programming language (NR)
- \$c Operating system (NR)

Notes

1. Subfield \$a contains the make and model of the machine on which the computer file is operated.
2. This field is a structured equivalent to field 542 (Mode of Use of Machine-Readable Data Files).

Examples

- 1 00\$aIBM PC\$bPascal\$cDOS 1.1
- 2 00\$aCompaq\$bBasic\$cDOS 3.2
- 3 00\$aApple II\$cDOS 3.3