

# **BIBLINK - LB 4034**

## **D8.1 User Requirements Specification**

*Name of Client:* European Commission

*Distribution List:* Pat Manson, European Commission  
Project Partners

*Author:* Robina Clayphan with  
contributions from Richard Carr,  
Rachel Heery, Andy Powell and  
Titia Van der Werf

---

*Authorised by:* Ross Bourne

---

*Contractual Date:* 28 February 1998

---

*Date of Issue:* 2 March 1998

---

*Issue:* 5

---

*Reference:* 5003/DEL-106

---

*Workpackage:* WP8 - Functional Specification

---

*Deliverable Number / Type / Nature* D8.1 / P / Specification

---

*Total Number of Pages:* 33

---

*Contact Details:* The British Library  
National Bibliographic Service  
Boston Spa  
Wetherby  
W. Yorkshire LS23 7BQ, UK

Telephone: 01937 546969  
Facsimile: 01937 546586  
Internet: [robina.clayphan@bl.uk](mailto:robina.clayphan@bl.uk)

---

**{TC ""\L}TABLE OF CONTENTS**

1. Document Control.....	1
1.1 Abstract.....	1
1.2 Keywords.....	1
1.3 Glossary.....	1
2. Management Overview.....	5
2.1 Executive Summary.....	5
2.2 Scope Statement.....	5
3. Introduction.....	6
3.1 Background to the project.....	6
3.1.1 Goal, Objectives and Benefits of the BIBLINK Project.....	6
3.1.2 Purpose of the System to be Developed.....	7
4. Summary of the Research Workpackages.....	8
4.1 Conceptual level.....	8
4.2 Scope (WP0).....	8
4.3 Metadata Formats and data set (WPs 1, 3 and 4).....	8
4.4 Identification of Publications (WP 2).....	9
4.5 Transmission of Data (WP 5).....	9
4.6 Authentication (WP6).....	9
4.7 Models to Accommodate Publisher Preferences (WP 3).....	9
5. Specification of User Requirements.....	11
5.1 Scope of the Software.....	11
5.2 Overview.....	11
5.2.1 Product Perspective.....	11
5.2.2 User Characteristics.....	12
5.3 The BIBLINK Working Environment.....	13
5.3.1 General Requirements.....	14
5.3.2 Operational Environment.....	14
6. Requirements for Events and Actions.....	15
7. Requirements for BIBLINK Workspace Records (BWR).....	16
7.1 Format Requirements.....	16
7.2 Mapping Requirements.....	16
7.3 Consistency requirements in the BWR.....	17
8. Requirements for the Creation of new records.....	18
9. Requirements for Editing records.....	19
10. Requirements for Searching.....	20
11. Requirements for Retrieving Records.....	21
12. Administration Requirements.....	22

13. Requirements For User Profiles..... 24

14. Interface Requirements ..... 25

15. Systems Requirements ..... 26

Annex A. BIBLINK Workspace Usage Scenarios ..... 27

    A.1 Scenario 1..... 27

    A.2 Scenario 2..... 28

    A.3 Scenario 3..... 29

## 1. DOCUMENT CONTROL

<i>Issue</i>	<i>Date of Issue</i>	<i>Comments</i>
1	18 November 1997	First draft circulated for comment
2	22 December 1997	Second draft circulated for comment
3	23 January 1998	Third draft circulated for acceptance
4	13 February 1998	Fourth draft circulated for comment
5	2 March 1998	Final version issued

### 1.1 Abstract

This document is the final version (5) of the User Requirements Specification for the BIBLINK demonstrator. It incorporates the comments received in response to the first draft considered at a review meeting on 8 December 1997, suggestions resulting from the issue of the second draft and a review by all partners of the third draft.

### 1.2 Keywords

BIBLINK, Library, bibliographic, national library, metadata, MARC, Dublin Core

### 1.3 Glossary

<b>Authentication</b>	BIBLINK has used this word to describe a process that will ensure a one-to-one relationship between a record and the publication it describes. In its accepted sense it is used in connection with identification of a user.
<b>BC</b>	see BIBLINK Core
<b>BD</b>	see BIBLINK Demonstrator
<b>BIBLINK Core</b>	the set of eighteen data elements defined for use in this project for the exchange of metadata between national bibliographic agencies and publishers. The element set is based on the Dublin Core.
<b>BIBLINK demonstrator</b>	a multi-national demonstrator, developed as part of the BIBLINK project, which provides an environment for the creation and editing of bibliographic records.
<b>BIBLINK Workspace</b>	Computer mediated co-operative working environment forming the core of the BIBLINK demonstrator for the creation and editing of bibliographic records for electronic publications.

<b>BIBLINK Workspace Record</b>	A group of bibliographic data elements held in different formats in the BIBLINK Workspace
<b>Bibliographic record</b>	a discrete bibliographic description stored either manually or electronically.
<b>BW</b>	see BIBLINK Workspace
<b>BWR</b>	see BIBLINK Workspace Record.
<b>CD-ROM</b>	Compact Disc Read Only Memory.
<b>DC</b>	see Dublin Core.
<b>DC-dot</b>	A Web based tool for generating Dublin Core metadata in a variety of formats.
<b>Deposit of publications</b>	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.
<b>DOI</b>	Digital Object Identifier
<b>DTD</b>	Document Type Definition.
<b>Dublin Core</b>	a metadata format defined on the basis of international consensus which defines a minimal information resource description for use in a WWW environment. The term 'Dublin' is used as Dublin, Ohio is the location of OCLC's headquarters.
<b>Electronic document</b>	see Electronic publication
<b>Electronic publication</b>	document, file, journal, etc. made available in electronic form.
<b>Electronic publisher</b>	see publisher.
<b>Format</b>	in the context of bibliographic control, the formalised structure in which the specific elements of bibliographic description are accommodated.
<b>HTML</b>	Hypertext Mark-up Language The standard language used for creating Web documents.
<b>HTTP</b>	HyperText Transfer Protocol. The protocol used for communication between Web clients and servers.
<b>Identification</b>	used in this document to refer to both a scheme to uniquely identify a document and a mechanism in the BW to identify the source of a record sent to the BW
<b>Integrity</b>	The unchanged state of the content and format of a message from sender to receiver.

<b>Internet Publisher</b>	an organisation or person who publishes documents on the Internet. These will be on-line documents.
<b>ISBN</b>	International Standard Book Number.
<b>ISSN</b>	International Standard Serial Number.
<b>Legal Deposit</b>	see Deposit of Publications.
<b>MARC</b>	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.
<b>Metadata</b>	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.
<b>National Bibliography</b>	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.
<b>NatMARC</b>	term used in this document to represent any national MARC format.
<b>NBA</b>	National Bibliographic Agency.
<b>Off-line publication</b>	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.
<b>On-line publication</b>	an electronic document which is bibliographically identifiable, which is stored in machine readable form on an electronic storage medium and is available on-line. For example a Web page.
<b>Participant</b>	A person or organisation taking part in the BIBLINK demonstrator and identified in the BW. These will include but not necessarily be limited to national bibliographic agencies, publishers and third parties such as identification agencies. See User.
<b>Partner</b>	The five national libraries and two academic institutions that comprise the BIBLINK consortium.
<b>Publisher</b>	an organisation or person who produces documents and makes them available.
<b>Serial</b>	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.
<b>SGML</b>	Standard Generalised Mark-up Language. ISO standard for document description separating content and structure.

<b>Third Party</b>	An organisation other than a publisher or national bibliographic agency who will have access to the BW in order to carry out specified operations on the BWR. For example, as an ISSN agency.
<b>UKMARC</b>	see MARC
<b>UNIMARC</b>	see MARC
<b>URL</b>	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: <a href="http://WWW2.echo.lu/libraries/en/libraries.html">http://WWW2.echo.lu/libraries/en/libraries.html</a>
<b>URN</b>	Uniform Resource Name
<b>USEMARCON</b>	Application developed as TAP project to convert between MARC formats using UNIMARC as an intermediate format.
<b>User</b>	A logical entity, associated with a Participant, for whom a user profile has been created to enable them to interact with the BW. Participants may be assigned more than one User identity for different operations in the BW.
<b>User Profile</b>	Set of criteria established for each User governing access rights and specifying those aspects of interaction with the BW that are configurable.

2.

## **MANAGEMENT OVERVIEW**

### **2.1 Executive Summary**

This report is the final version (5) of the User Requirements Specification for the system to be developed for the BIBLINK project. Section 3 outlines the background to the project. The requirements as they were expressed in each of the research workpackages in the first stage of the project are summarised in Section 4. The remainder of the report specifies the user requirements and has been drawn from the preceding drafts of this report which have been reviewed by all the project partners.

### **2.2 Scope Statement**

Based on the research and recommendations from the work packages undertaken in Stage 1 of the BIBLINK project this specification itemises the user requirements for the BIBLINK Workspace which will form the central component of the demonstration system to be developed and implemented in the second stage of the project. The full demonstration system consists of operations that will be carried out at various locations and points in time by the publishers, the NBAs, third parties as well as the central workspace. This URS forms the basis for the Functional Specification.

3.

## INTRODUCTION

### 3.1 Background to the project

There is continuous growth in the amount of material being published chiefly or solely electronically. This is a cause for concern for national libraries as there is the danger of significant works being omitted from the national bibliography and other services. In addition there is no adequate bibliographic control over electronic publications nor is there an agreed standard of bibliographic description for many types of electronic publication which might assist with bibliographic control. Finally, there is at present no direct bibliographic link between publishers and national bibliographic agencies.

#### 3.1.1 *Goal, Objectives and Benefits of the BIBLINK Project*

The overall goal of the BIBLINK project is to further the development and improvement of national bibliographic services.

The objectives are:

- to create a direct bibliographic link between publishers and national bibliographic agencies which will facilitate adequate bibliographic control over electronic publications, especially those which are published via the networks.
- to develop an agreed bibliographic description for many types of electronic publication which might assist with bibliographic control.
- to meet the needs of the national libraries to secure authoritative advance information about new electronic publications
- to exercise bibliographic control over those publications
- to enable publishers of electronic materials to register new electronic publications
- to enable publications to carry enriched bibliographic data as an integral part of the electronic publication
- to help second level users such as library utilities to meet an increasing demand for information about electronic publications
- to be able to identify those publications uniquely (particularly important for material published via the networks).
- to supply third level users - libraries - with information about electronic publishing in order to be able to satisfy their own users
- to supply such libraries with bibliographic records relating to this material for integration into their own catalogues.

Many of the benefits of such a system are implicit in the above objectives. The main benefit will be to avoid the danger of significant works being missed by the national libraries due to being produced only in electronic form.

### **3.1.2 Purpose of the System to be Developed**

The needs of the national libraries are to secure authoritative information about new electronic publications and to exercise bibliographic control over those publications. For the publishers of electronic materials there is a need to register new electronic publications and to make those publications accessible to potential users both by inclusion in the national bibliography and by carrying bibliographic metadata.

The solution is to develop a system that will form the central part of a set of procedures allowing publishers to submit metadata for use by NBAs. This data can be enriched by third parties and enhanced and converted into national MARC formats for use in the national bibliography, the resulting data being made available to the publishers for their own uses, where possible. Not all of these activities will be performed by the piece of software being specified in this document.

The system will be sufficiently flexible to allow different NBAs to implement procedures that will suit their own local practices. For example, some may wish to use the enhanced metadata in the national bibliography before seeing the publication (akin to a CIP operation) whilst others may prefer to create a full record only after the publication has been seen.

This URS is intended to be read by the participants in the project and the developers of the software. It is to be used as the basis for the production of the Functional Specification and as the specification against which to assess the resulting software.

4.

## SUMMARY OF THE RESEARCH WORKPACKAGES

The first eighteen months of the project were spent researching the components of the BIBLINK Demonstrator. As a result of this research various conclusions were reached. The results of the research workpackages are summarised in the following sub-sections to provide background information.

### 4.1 Conceptual level

A metadata transmission and conversion system is required that will allow data relating to electronic publications, both on-line and off-line, to be sent by publishers to national bibliographic agencies for inclusion in the national bibliography. The system will provide for the conversion of the metadata into a format acceptable to the NBAs, the enhancement of the metadata to the standards required by the libraries according to local practices and its return to the originating publisher where possible.

### 4.2 Scope (WP0)

This work package defined the range of electronic publishers and electronic publications to be considered by the project. The scope includes all those publications that would traditionally be included in a national bibliography in whatever electronic medium they are published. A working definition of "document-like objects" was proposed which does not limit the scope to text-only documents<sup>1</sup>. A comprehensive list of publisher types covering the established and newly emerging sectors, was also produced from which publishing partners would be recruited at a later point in Stage 1 of the project.

### 4.3 Metadata Formats and data set (WPs 1, 3 and 4)

A simple metadata set has been identified consisting of the following 18 data elements<sup>2</sup>:

Author	Keywords	System requirements	Contributor
Title	Description	Format	Identifier
Publisher	Edition/version statement	Language	Frequency
Price	Date of publication	Terms and conditions	Checksum
Extent (Size)	Place of Publication		

These have been mapped to the Dublin Core<sup>3</sup> and are referred to in the remainder of this document as the BIBLINK Core (BC). The syntax required to express the elements in HTML META tags has been examined and it is agreed that this will be suitable for BIBLINK<sup>4</sup>. The syntax must be finalised and the format for its transmission must be defined. It is anticipated that an ASCII file format will be used.

As it is the international exchange format, UNIMARC will be used as the intermediate format in order to make BIBLINK records widely available and ensure the adaptability of BIBLINK to other users.

It is considered important to build on existing work in the libraries sector and use and, where necessary, develop applications that are already available. To this end it is recommended that where conversions between NatMARC and UNIMARC are required USEMARCON should be used to perform conversions and extended

---

<sup>1</sup> D0.1- Scoping Document, 5003/Del/21, October 1996

<sup>2</sup> D3.1 Minimum Dataset, 5003/Del-86, September 1997.

<sup>3</sup> D4.1 - Format Conversion, 5003/Del-85, September 1997

<sup>4</sup> Ibid

to accommodate those MARCs not yet implemented in it. It is recognised that this would entail timely and accurate input of expertise from the NBAs to enable the developer to complete the work.

#### 4.4 Identification of Publications (WP 2)

It was recognised that unless national bibliographic agencies choose to take an active role in the attribution of an identifier, such as the DOI or the URN (possibly using national bibliography numbers), then the BW would simply accept the identifiers that are supplied by the publishers. The system should accept the addition of an identification number during enhancement by the NBA or by a third party and handle the return of that identifier to the publisher.

#### 4.5 Transmission of Data (WP 5)

The data will be transmitted using internet protocols that are widely available to the parties concerned. Research carried out amongst the publishing communities at a later date (WP3) narrowed this down to email and Web technologies. Depending on project resources it was thought that it might be possible to develop an NBA Web crawler.

For security, the data transmission mechanism should allow the receiver to authenticate the sender in some way. The system should make provision for an audit trail tracking the progress of each record through the system and back to the originator. The volume of data is not expected to be great at any one time, at most being a small number of records at once. In terms of frequency of transmission, the system does not need to be interactive, records will be sent between the parties intermittently. The chosen transmission solution will need to support the character sets of the European languages participating as well as binary data. These are Catalan, Dutch, English, French, Norwegian and Spanish.

#### 4.6 Authentication (WP6)

The issue of authenticating the one-to-one relationship of a record and the publication it describes will be handled by means of including a hash value in the data set. This value can be generated and entered in the record by either the publisher or the national bibliographic agency by agreement between the communicating parties.

#### 4.7 Models to Accommodate Publisher Preferences (WP 3)

Electronic publishers are not a homogenous group<sup>5</sup> so different methods must be provided to allow them to contribute metadata. For the purposes of the BD the principal difference is the quantity of metadata held per publication. Certain of the publishers have comprehensive databases from which they can generate the required elements in the required format, others do not have these facilities and will need a tool to facilitate creation of the data. It is envisaged that this will be similar to the DC-dot<sup>6</sup> generator. The third method will be of use to publishers who use a standard SGML DTD who can transmit that to the NBA for conversion to a richer record than can be created from the simple metadata set.

WP3 did not support the idea of using EDI for packaging the data and elected to use a simple formatted text file. This does not mean the BD should not adopt a "packaging" approach if it seems appropriate when transmitting data in more than one format e.g. NatMARC and BC.

---

<sup>5</sup> D5.1 - Transmission of Data, 5003/Del-58, February 1997 and D3.2 - Consensus Building, 5003/Del-87, August 1997.

<sup>6</sup> <http://www.ukoln.ac.uk/metadata/dcdot/>

Email and Web technology were nominated as the preferred transmission methods. Given that no one publisher will transmit a large amount of data at any one time, using email is the preferred option. It must be possible for records to be sent singly or in groups.

The publishers had different views as to the sort of record (if any) they wanted to receive in return. Several are interested in receiving one or two of the data elements which the NBA would have modified and at least one may be able to make use of the full national MARC record. The system must therefore be able to return either the enhanced version of the BC from which the publishers can extract the elements they want or the enhanced national MARC record. Local agreements will be made between NBAs and publishers about the timescales for production of the enhanced data so the BD mechanism for monitoring records must take account of these timescales.

5.

## **SPECIFICATION OF USER REQUIREMENTS**

The following sections are based on the conclusions and decisions expressed in Section 4 of this report and a review by BIBLINK partners of the preceding drafts of this document.

Specific requirements have been labelled and numbered to facilitate reference. They have also been designated as mandatory, highly desirable or desirable. This is indicated in the right hand column of the table by the numbers 1, 2, or 3 respectively.

English has been adopted as the working language of the project and this will apply to communications and documentation during the development of the system. Further reference to the languages the system should support can be found in specific requirements in the following sections.

### **5.1 Scope of the Software**

The proposed software, referred to as the BIBLINK Workspace (BW), will meet the requirements of the users expressed in this document (URS), which have been drawn from BIBLINK research work, and be implemented according to the Functional Specification based upon the URS. The BW will provide a mechanism to allow the deposit of metadata and the subsequent retrieval, editing and updating of that data by various parties.

### **5.2 Overview**

#### **5.2.1 Product Perspective**

National libraries receive publications through the (legal) deposit system and create or derive MARC records for them for use in the national bibliography and other services. These other services include the supply of records to other libraries, information providers and intermediaries. Each nation has developed a version of MARC that is best suited to its own cataloguing requirements and associated exchange formats which, for the most part, are ISO2709 compatible. Many of the variations of the MARC format are closely based on either UK or US MARC.

To facilitate the international exchange of records UNIMARC has been (and continues to be) developed as a switching mechanism. In theory each nation need only develop a conversion programme that will convert records from their own MARC format to UNIMARC and back to be able to use the records from any other nation. The USEMARCON project implemented the use of UNIMARC as an exchange mechanism by creating the mapping tables and rules files for a limited number of national MARCs.

Imported and exported records are transported using a wide range of technologies: FTP, email, tape, CD-ROM, Web technology and Z39.50 applications.

Publishers also use a wide range of technologies in the production and marketing of their publications. The production side tends to be a separate operation from marketing and it is the latter that has most comprehensively adopted IT for communications. The production operation is the area that holds the bibliographic data required for BIBLINK and here the use of technology tends to be more limited and there is very little standardisation. The larger publishers hold comprehensive information relating to publications in a format appropriate to their own functions. The smaller publishers, including many of those publishing on the Internet, tend not to produce comprehensive bibliographic information. All those consulted used email and the Web for communication and information dissemination and many made their products available on the Web as well.

### **5.2.2 User Characteristics**

The immediate users of the BW will be the publishers who will be participating in the development and demonstration of the system, the project partners, who comprise five national libraries and two academic institutions, and third parties who may be involved in adding or amending data to records. In the long term it is hoped that many more publishers will avail themselves of the system to submit data to NBAs and that additional NBAs will adopt the system as well.

The BW should offer a choice of interfaces to users - this is especially important in the case of the publishers who may not be willing to participate in a system that is not immediately accessible. Similarly, the means of interacting with BW should be as simple as possible so as not to require a high level of technical expertise on the part of the publishers in particular.

#### **5.2.2.1 The Publishers**

The publishers do not form a homogenous group in terms of IT use or available technical expertise and will therefore require a range of methods by which to interact with the BW.

Publishers opting to generate the minimum data set from their existing systems will need the programming expertise to generate the record and present the data in the agreed format according to specifications documented for the BW. They will need the technical ability to interact with the BW to create a new record and access it at a later date to download enhanced data.

Publishers without a comprehensive bibliographic database may want to use an interface that prompts for entries in each element of the record being created. A user should need only basic IT abilities in order to use the Web and operate this BW interface. A user manual should be made available for this. No programming experience will be needed in this instance.

Publishers wishing to submit data using SGML will need knowledge of their own system in order to be able to identify and transmit the data via the BW interface.

In all cases users will need a measure of technical expertise to handle the data received in the return transmission. This lies outside the domain of the BW.

#### **5.2.2.2 Third Parties**

As currently envisaged, the role of the third parties will be to add one or more data elements to records in the BW. They will need sufficient technical ability to access and edit the necessary fields.

#### **5.2.2.3 The National Bibliographic Agencies**

Users at the NBAs will need: sufficient knowledge of their own systems and MARC formats and cataloguing rules and procedures to do all or some of the following:

- be able to handle UNIMARC records
- be able to handle BC records in HTML
- convert these into their own national MARC formats
- enhance the data into a MARC record of an acceptable standard for inclusion in the national bibliography

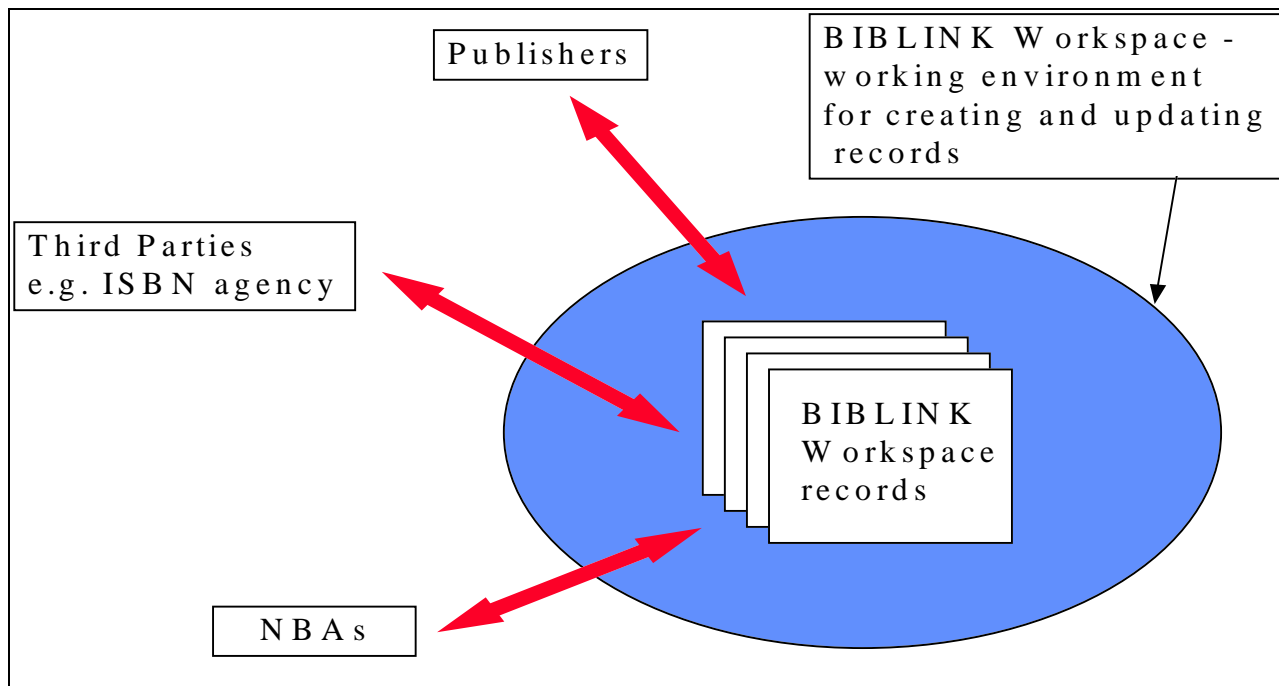
- add a NatMARC record to the BW
- add a BC record to the BW

### 5.3 The BIBLINK Working Environment

The core of the BIBLINK demonstrator will consist of a computer mediated working environment called the BIBLINK Workspace (BW). It can be envisaged as a virtual workspace encompassing a database and functionality that will, in the first instance, allow publishers to create records and subsequently give access to participants to retrieve, update and, ultimately, delete those records. The BW will convert the data between the formats required by the various parties and provide the mechanisms necessary for the functioning and management of the system. It will allow the various parties to view and download the records or elements in the records in different formats at various stages in the development of the record. Users will work in the BW according to a pre-established user profile defining access rights and those aspects of interaction with the BW that are configurable.

The NBAs will vary in the way they apply the use of the BW to their current procedures and each will specify the flow of work within their own organisation. They will want to configure the actions that the BW carries out in relation to events in the life of the BWR according to local practices. For example one NBA may want to be notified about a new record as soon as it is created by a publisher, whereas another will only want to be notified once a specified third party has added an identifier to a new record. The situation will be similar with the publisher participants: one may want to be sent an enhanced BC record as soon as the NBA has updated the original but another may wish only to be notified that an identifier has been added to allow them to retrieve it when they are ready to do so. These preferences will be specified in the User Profile established for each participant. To give a picture of how the flow of metadata from publisher to NBA and back is envisaged, two examples of usage scenarios are given in Annex A. A candidate list of events and actions is given in Section 6.

The diagram below outlines a model of the BIBLINK Workspace.



Although the BW is a demonstration system it is anticipated that it will be deployed on a wider scale after the demonstration has finished.

### 5.3.1 General Requirements

General 1. MARC formats regularly undergo change and it may be necessary to modify the software in the BW to accommodate such changes. Tools must be made available to allow changes to be made to any mapping tables or conversion software supplied as part of the BW. Where possible existing tools that are in use by, or are familiar to, all or some of the partners, such as USEMARCON, should be preferred to the development of customised solutions that will entail any additional commitment for future maintenance.	1
General 2. The solutions supplied must be sufficiently robust and extensible to stand the test of time. The use of an application or service that may not be available in the foreseeable future should be avoided.	1
General 3. As the users of the BW will be a diverse group in every respect it is important that the solution offered is hospitable to as wide a range as possible. Where the developer has a choice of methods of implementation the BIBLINK partners will be consulted as to the preferred solution.	1
General 4. Publishers will either create records directly by entering data via a BW interface or by transmitting data harvested from their own databases. Similarly, any user wanting to copy a record for use in their own database will retrieve the record via the BW interface for loading into their own system. The software will not therefore need to interface directly with any publisher, NBA or third party system.	1
General 5. User Guides must be provided for participants and users: the partners must be consulted about the text of user documentation.	1
General 6. The functionality defined for the demonstrator must include the return of data to publishers which is regarded as equally important as delivery of the data to the NBAs.	1

### 5.3.2 Operational Environment

The BW will form the core component of the BIBLINK demonstrator. It is envisaged that multiple BWs will be run for the demonstrator - one at each participating NBA. Where an NBA is unable to host their own BW, a central one will be provided. It is likely that this will be hosted by UKOLN<sup>7</sup>.

6.

---

<sup>7</sup> UKOLN, the United Kingdom Office for Library and Information Networking, is a national centre for support in network information management in the library and information communities. It provides awareness, research and information services and is a member of the BIBLINK consortium

## REQUIREMENTS FOR EVENTS AND ACTIONS

Section 5.3 referred to participants wishing to configure events and actions in the BW to suit their local practices. The following is a set of candidate Events and Actions that must be included in the BW. They need not be limited to those listed as it is recognised that some actions will be performed as part of the internal system processes of the BW and will be hidden from the user. In addition some events will be configurable by the administrator, for example a purge action to delete records triggered by the age of a record.

Event 1. Record created	1
Event 2. Record deleted	1
Event 3. Field modified by <user>	1
Event 4. Specified age of record reached	1

Action 1. Notify user of event	1
Action 2. Transmit record with notification	1
Action 3. Update field	1
Action 4. Mark field as out of date	1
Action 5. Purge records	1

7.

## REQUIREMENTS FOR BIBLINK WORKSPACE RECORDS (BWR)

The BIBLINK Workspace Record is envisaged as a data structure that can accommodate the data elements in the formats that will be supported by the BW. These are the BIBLINK Core, SSSH, SGML non-serial header, UNIMARC records, NatMARC records.

### 7.1 Format Requirements

Different types of user of the BW will want to see different representations of the data contained in a BWR. A dataset of eighteen elements based on the Dublin Core has been defined by for use in the demonstrator. Some publishers will send all or some of these elements to the BW to create a BC record. The fields in the BC are repeatable so the data set may contain more than eighteen fields. In addition, fields in the BC may have an associated SCHEME, indicating that the value is taken from an externally maintained controlled list (e.g.' UDC) or that it is formatted according to an external specification (e.g. ISO-8601) Other publishers may prefer to send an SGML header for conversion to a BC record. The library community will want UNIMARC and NatMARC records.

Format 1. The BWR will have the ability to represent a BC record including any repeated fields and where necessary associating a SCHEME with each field.	1
Format 2. The BWR will have the ability to represent an SGML header (both SSSH and non-serial header).	1
Format 3. The BWR will have the ability to represent a full or partial UNIMARC record.	1
Format 4. The BWR will have the ability to represent a full or partial NatMARC record.	1

### 7.2 Mapping Requirements

Map 1. All data received in SGML header format will be mapped to BC format where the publisher has supplied the information necessary to produce a mapping between the two formats.	1
Map 2. All data in BC format will be mapped to UNIMARC format.	1
Map 3. Data in UNIMARC format will be mapped to NatMARC format if the NBA responsible for the NatMARC has supplied the mappings from UNIMARC to NatMARC.	1
Map 4. Data in NatMARC format created by NBAs will be mapped to UNIMARC format where the NBA responsible for the NatMARC has supplied mappings from NatMARC to UNIMARC.	1
Map 5. Updated data in UNIMARC format will be mapped back to BC format.	1

### 7.3 Consistency requirements in the BWR

Records in the BW will be dynamic in that various parties will edit them at different points in time. It is important that the "authoritative" version of a record that has been enhanced by an NBA or third party should not be overwritten by updating carried out by a publisher.

*D8.1 User Requirements Specification*

Map 6. The BW will support the updating of fields in an existing BWR and ensure that updates are flagged as such.	1
Map 7. The BW will protect authoritative data from being overwritten.	1
Map 8. The BW will provide a mechanism to ensure that any publisher updates to an existing record are flagged for the attention of NBAs and third parties.	1
Map 9. The BW will institute a version control mechanism on the changing BWR.	1

8.

## REQUIREMENTS FOR THE CREATION OF NEW RECORDS

New BWRs will be created by publishers supplying at least the 'title' and 'publisher' elements of the record or as many as configured as mandatory by the administrator. Some publishers will wish to generate the BC elements from their existing databases and deposit this in the BW. Others will require a tool that will prompt for entries in each element of the BC record and allow the deposit of the resulting data set. A third group will wish to create a record by depositing an SGML header. Publishers in the third group must supply a DTD to enable the mapping of the contents to the BC elements. Of the transmission options considered, publishers will be able to elect to use either email or Web technology.

New 1. Publishers are the only parties who will be permitted to create a new BWR.	1
New 2. The BW will provide a mechanism to handle new BWRs not conforming to the mandatory elements.	1
New 3. The BW will provide interfaces to allow the transmission of a full or partial BC record generated from a publisher's database using either email or Web technology.	1
New 4. The BW will provide interfaces that will prompt a publisher to provide the BC elements and add the resulting record to the BW.	1
New 5. The BW will provide interfaces to allow publishers to send an SGML header using email or Web technology.	1
New 6. It is highly desirable that the BW provides a mechanism to support the creation of multiple new records when multiple records are sent by publishers in one transmission.	2
New 7. Accessing a publication - as part of the record creation process the BW will provide a mechanism that will allow publishers to provide additional information about how the NBA may access an on-line publication to assist with record enhancement.	1
New 8. The BW will at the minimum support data entry in the character sets of the participating parties (ISO 8859). To be extensible it is highly desirable that support should be offered for non-Latin characters as well, e.g. Unicode.	1 2
New 9. The BW will provide interfaces that will allow the transmission of BC elements in HTML.	1

9.

## REQUIREMENTS FOR EDITING RECORDS

Once a BWR has been initiated by a publisher, NBAs and third parties will access it to add or amend data.

NBAs will vary as to the format of the record they wish to see in order to carry out their own enhancements upon it for inclusion in the National Bibliography. At different points in time they may wish to access the UNIMARC record and/or the BC and/or the SGML header and/or the NatMARC record.

Local practices will determine the point at which the NBAs wish to access the record. The sequence in which other parties access the record to carry out their different functions will vary from one country to another.

Third parties may be involved in adding or modifying data elements supplied by publishers. For example, an NBA may arrange for an ISBN agency to add an ISBN to the BC where the original record is created without an identifier.

For the purposes of the demonstrator the records will persist for the duration of the demonstrator or until a user or the administrator deletes a record.

Edit 1. The BW will provide a choice of mechanisms for receiving notification about the creation of a new record using email or Web technology.	1
Edit 2. The BW will provide the interfaces for users to see and edit a record using either email or Web technology.	1
Edit 3. The BW will provide a mechanism to allow users to customise the view of the BWR they wish to see and edit.	1
Edit 4. The BW will provide a choice of mechanisms for receiving notification of events in the life of a BWR using email or Web technology.	1
Edit 5. The BW will support local variations as to the sequence in which users access a BWR.	1
Edit 6. The BW will provide a mechanism for limiting access of users to the elements of a record relevant to their functions.	1
Edit 7. The BW will provide a mechanism to support the deletion of records.	1
Edit 8. The BW will provide a mechanism to enable an authorised user to designate a record as "finished" which would prevent any further updating.	1

10.

## REQUIREMENTS FOR SEARCHING

Users will need to be able to search the records in the BW.

Search 1. The records in the BW must be searchable using the Web interface.	1
Search 2. Fielded Boolean searching across all or some fields is preferred	2
Search 3. Simple text searching	3
Search 4. It is desirable that a search function is also available in the email interface.	3

11.

### REQUIREMENTS FOR RETRIEVING RECORDS

All parties may want to access a record, or certain elements of a record, in order to retrieve all or part of it for use in their own systems or publications.

Retrieve 1. The BW will provide a mechanism to enable the retrieval of a record in UNIMARC format using either email or Web technology.	1
Retrieve 2. . The BW will provide a mechanism to enable the retrieval of record in NatMARC format using either email or Web technology.	1
Retrieve 3. The BW will provide a mechanism to enable the retrieval of an SGML header using either email or Web technology.	1
Retrieve 4. The BW will provide a mechanism to enable the retrieval of a tagged HTML BC record (HTML 3.2 and HTML 4.0) using either email or Web technology.	1
Retrieve 5. The BW will provide a mechanism to enable the retrieval of a plain text BC record using either email or Web technology.	1
Retrieve 6. The BW will provide a mechanism to enable the retrieval of selected fields of a BWR using either email or Web technology.	1

12.

## ADMINISTRATION REQUIREMENTS

The NBAs will have overall control of the records in the BW and the access rights assigned to each participant and user. Participants are defined as the three bodies taking part in the BD, that is Publishers, NBAs and nominated third parties such identification agencies (e.g. ISSN agencies). A user is defined as an entity associated with a participant adopting a specific role to interact with the BW and having a specific user profile for that purpose. For example, a publisher participant may want more than one user profile for the creation of records for different types of publication, or an NBA participant may want a user profile to create records in the role of publisher of its own publications.

Admin 1. The BW will support the functions of a BW Administrator.	1
Admin 2. . The BW will provide a mechanism to create, modify and delete Participants.	1
Admin 3. The BW will provide a mechanism to hold a directory entry for each Participant containing contact details.	1
Admin 4. The BW will provide a mechanism to create modify and delete Users.	1
Admin 5. The BW will provide a mechanism to associate Participants with Users.	1
Admin 6. The BW will provide a mechanism to create a User Profile and assign it to a User.	1
Admin 7. The BW will provide a mechanism to associate actions to events.	1
Admin 8. The BW will provide a mechanism to configure those actions and events that are configurable by a User.	1
Admin 9. The BW will provide a mechanism to configure which fields in the BWR are mandatory given an absolute minimum of "Title" and "Publisher".	1

The administration of the system will also necessitate the following:

Admin 10. Identification of users - the BW will include a mechanism to verify the identity of the users interacting with the BW.	1
Admin 11. Audit trail - the BW will include a mechanism to register the creation of a new BWR and use this to monitor its progress/history in the BIBLINK demonstrator.	1
Admin 12. Rejection of record - the BW will provide a mechanism for sending a message of rejection to publishers whose publication the NBA has declined to accept for inclusion in the national bibliography.	1
Admin 13. Rejection messages - the BW will provide a mechanism for sending other rejection messages to users.	1

Admin 14. Database maintenance - the BW will provide mechanisms for the management and maintenance of the database. These should include periodic reports on database activities and numbers of records, and backup and recovery procedures..	1
Admin 15. The BW will incorporate mechanisms to validate elements in the BWR in accordance with the Functional Specification	1

13.

### REQUIREMENTS FOR USER PROFILES

The requirements in the preceding sections indicate many points where Users can configure various aspects of the BW to suit their particular requirements. Access to the BW will be governed by a User Profile assigned to each User.

Working within a set of rules for operations in the BW and under the control of the BW Administrator the User Profile should allow the user to configure the following:

Profile 1. The transfer mechanism: email or Web form	1
Profile 2. The syntax for communication	1
Profile 3. The format of the records to be created	1
Profile 4. Actions associated with events. ( e.g. Event: NatMARC record created from BC, Action: email NatMARC record to NBA.)	1
Profile 5. Access rights per field and per record e.g. read/write/delete	1
Profile 6. The record format the user wishes to view e.g. BC, NatMARC	1
Profile 7. The record elements the user wishes to view	1
Profile 8. The format of record to be downloaded - plain text, HTML tagged, NatMARC, etc.	1
Profile 9. The language of the interaction with the BW.	1
Profile 10. The User Profile will indicate the Participant associated with the User.	1

14.

## INTERFACE REQUIREMENTS

Interface 1. It is highly desirable that all text associated with the interface be available in the languages of the partners: Catalan, Dutch, English, French, Norwegian, Spanish.	2
Interface 2. The BW should support the character sets of all participating partners - Catalan, Dutch, English, French, Norwegian, Spanish. As it is hoped that BIBLINK will be extensible to other participants in the future it would be preferable if the BW had the potential to support non-Latin based character sets as well.	2 3
Interface 3. The BW must use industry standard applications for the interface such as Netscape Navigator and MS Internet Explorer.	1
Interface 4. The BW interfaces must be clear and immediately understandable.	1
Interface 5. An appropriate "Help" facility must be available at all times in Web interface which should include specific help on each field in the BW record. It is also desirable that that it is available in the email interface.	1 3
Interface 6. The partners must be consulted about the design of the interface and the wording for various messages to the user.	1

15.

## SYSTEMS REQUIREMENTS

It is of the utmost importance that the BW should not only be usable now but be capable of extension and maintenance in the future.

Systems 1.	Industry standard hardware and software platforms will be used as specified in the Functional Specification.	1
Systems 2.	The BW server will run on hardware and software platforms that will provide an acceptable level of performance for multi-user access.	1
Systems 3.	The BW will be portable with minimal modifications to different platforms.	1
Systems 4.	For the purposes of the demonstrator a minimum of 2000 BWRs will be created and stored in the BW. Subsequent implementations will result in this number being considerably greater. The BW should therefore be capable of expansion.	1

## **ANNEX A000. BIBLINK WORKSPACE USAGE SCENARIOS**

Andy Powell

<http://www.ukoln.ac.uk/metadata/biblink/wp8/usage-scenarios/>

The following three scenarios show examples of how metadata can be shared between publishers, NBAs and third parties using the BW. These examples are included to show possible patterns for the flow of data between participants and are not intended to be comprehensive. At various points in the flow, BW Events occur. An Event may be a record being created, a field in a BWR being updated or something else. Associated with each Event will be one or more Actions. An Action may be to send a particular view of a BWR somewhere, to update a field in a BWR or something else. The Actions that will be triggered in the BW by the Events described here are detailed in the functional specification.

### **A.10 Scenario 1**

In this scenario we consider a small Internet publisher Publisher-X, who wish to supply simple metadata about their publications to the National Bibliographic Agency (NBA) in order that each publication appears in the National Bibliography and so that an enhanced Dublin Core-like bibliographic record can be embedded into the HTML of Publisher-X's Web pages.

Publications are made available on the Web prior to notification being sent to the NBA. No formal 'identifier', such as an ISSN or DOI, is required for each publication. The NBA in question has provided a set of mapping tables to allow the minimal BIBLINK UNIMARC record to be mapped to and from NatMARC records.

Here is a brief description of the flow of metadata between Publisher-X and the NBA using the BIBLINK Workspace (BW).

1. Publisher-X makes the Web pages for a new publication available on their Web server in the normal way.
2. The publisher uses the BW Web form to create a new BIBLINK Workspace Record (BWR). This new record can be as simple or as complex as required (using the 18 elements in the BIBLINK Core (BC)) but must contain all the elements that the Administrator has configured as mandatory in the BW (typically 'Title' 'Publisher' and 'Identifier').
3. The BW converts the BC into a UNIMARC record and stores it into the UNIMARC field in the new BWR.
4. The BW converts the UNIMARC record into a NatMARC record, and stores it into the NatMARC field in the BWR.
5. The BW sends an email message to the NBA containing a text copy of the BC and the NatMARC record.
6. The NBA loads the NatMARC record into their local MARC based database and enhance it, viewing the Publisher's Web pages if necessary.
7. The NBA uses email to send the enhanced version of the NatMARC record back to the BW.
8. The BW stores the enhanced NatMARC record into the NatMARC field of the BWR.
9. The BW converts the enhanced NatMARC record into an enhanced UNIMARC record using the minimal mapping supplied by the NBA.
10. The BW updates some or all of the BC fields based on the enhanced UNIMARC record.

11. The BW sends an email message to Publisher-X containing an HTML (META tag) version of the enhanced BC fields.
12. The NBA sends an email message to the BW indicating that the BWR is now complete. (Note that it may well be possible to combine this with step 7 above).
13. The BW locks the BWR so that no further updates can be made to it.
14. Publisher-X copies the HTML META tags from the email message and embeds them into the HEAD section of the publication's home page.

## **A.20 Scenario 2**

In this scenario we consider a medium sized Internet publisher, Publisher-Y, who use SGML to prepare their publications and who wish to supply this to the National Bibliographic Agency (NBA) in order that each publication appears in the National Bibliography and so that an enhanced version of the SGML header can be obtained.

Publications are made available on the Web some time after notification has been sent to the NBA. An ISSN is required for each publication. The NBA in question has provided a set of mapping tables to allow the minimal BIBLINK UNIMARC record to be mapped to and from NatMARC records.

Here is a brief description of the flow of metadata between Publisher-Y and the NBA using the BIBLINK Workspace (BW).

1. Publisher-Y is working on a new publication and has prepared a minimal SGML header to describe it.
2. Publisher-Y sends a copy of the SGML header to the BW using email.
3. The BW creates a new BWR and stores the SGML header in the SGML-1 field.
4. The BW converts SGML-1 into as many BC fields as necessary.
5. The BW converts the BC into a UNIMARC record and stores it into the UNIMARC field in the new BWR.
6. The BW converts the UNIMARC record into a NatMARC record, and stores it into the NatMARC field in the BWR.
7. The BW sends an email message to the ISSN Agency containing a text copy of the BC.
8. The ISSN Agency assigns a new ISSN to the publication and sends an email message back to the BW to update the BC Identifier field (note this will actually cause the addition of a second BC Identifier field). The BC Identifier field is the only field in the BWR that the ISSN has been given access to modify.
9. The BW converts the BC into a UNIMARC record and stores it into the UNIMARC field in the BWR. It also flags the existing SGML-1 record as being out of date.
10. The BW converts the UNIMARC record into a NatMARC record, and stores it into the NatMARC field in the BWR.
11. The BW sends an email message to the NBA containing a text version of the DC fields.

12. The NBA notes the new record but doesn't proceed until the publication appears on the Web.
13. When the new publication is made available on the Web the NBA uses the Web interface to the BW to retrieve the NatMARC record and SGML-1 header.
14. The NBA loads the NatMARC record into the local MARC based database and uses the SGML-1 header and the Web publication to enhance the NatMARC record.
15. The NBA uses the Web interface to the BW to upload the enhanced version of the NatMARC record back to the BW.
16. The BW stores the enhanced NatMARC record into the NatMARC field of the BWR.
17. The BW converts the enhanced NatMARC record into an enhanced UNIMARC record using the minimal mapping supplied by the NBA.
18. The BW updates some or all of the BC fields based on the enhanced UNIMARC record.
19. The BW sends an email message to Publisher-Y containing a text version of the enhanced BC fields.
20. The NBA sends an email message to the BW indicating that the BWR is now complete. (Note that it may well be possible to combine this with step 15 above).
21. The BW locks the BWR so that no further updates can be made to it.
22. Publisher-Y uses the text version of the enhanced BC fields to create an enhanced SGML header for internal use.

### **A.30 Scenario 3**

In this scenario we consider a CD-ROM publisher Publisher-Z, who wish to supply simple metadata about their publications to the National Bibliographic Agency (NBA) in order that each publication appears in the National Bibliography and so that an enhanced Dublin Core-like bibliographic record can be embedded into an "electronic title page" of Publisher-Z's CD-ROM.

Publisher Z has a comprehensive database containing many data elements pertaining to each publication - these include most of the BIBLINK Core elements.

Notification is sent to the NBA well ahead of publication. A formal 'identifier', such as an ISBN or DOI, is required for each publication. The NBA in question has provided a set of mapping tables to allow the minimal BIBLINK UNIMARC record to be mapped to and from NatMARC records.

Here is a brief description of the flow of metadata between Publisher-Z and the NBA using the BIBLINK Workspace (BW).

1. Publisher-Z is planning the publication of a CD-ROM and wishes to include an "electronic title page" containing a Dublin Core-like bibliographic record.
2. The publisher generates as many of the BC metadata elements as possible from their database in the format specified for the BW and send the resulting record to the BW as an email message.

3. The BW converts the BC into a UNIMARC record and stores it into the UNIMARC field in the new BWR.
4. The BW converts the UNIMARC record into a NatMARC record, and stores it into the NatMARC field in the BWR.
5. The BW sends an email message to the ISBN Agency containing a text copy of the BC.
6. The ISBN Agency assigns a new ISBN to the publication and sends an email message back to the BW to update the BC Identifier field (note this will actually cause the addition of a second BC Identifier field). The BC Identifier field is the only field in the BWR that the ISBN agency has been given access to modify.
7. The BW converts the BC into a UNIMARC record and stores it into the UNIMARC field in the BWR.
8. The BW converts the UNIMARC record into a NatMARC record, and stores it into the NatMARC field in the BWR.
9. The BW sends an email message to the NBA containing a text copy of the BC and the NatMARC record.
10. The NBA loads the NatMARC record into their local MARC based database and enhance it.
11. The NBA uses email to send the enhanced version of the NatMARC record back to the BW.
12. The BW stores the enhanced NatMARC record into the NatMARC field of the BWR.
13. The BW converts the enhanced NatMARC record into an enhanced UNIMARC record using the minimal mapping supplied by the NBA.
14. The BW updates some or all of the BC fields based on the enhanced UNIMARC record.
15. The BW sends an email message to Publisher-Z containing a text version of the enhanced BC fields.
16. The NBA sends an email message to the BW indicating that the BWR is now complete. (Note that it may well be possible to combine this with step 11 above).
17. The BW locks the BWR so that no further updates can be made to it.
18. Publisher-Z copies the BC record from the email message and embeds it into the "title page" of the publication.