Network Working Group
Request for Comments: 3187
Category: Informational

J. Hakala
Helsinki University Library
H. Walravens
The International ISBN Agency

October 2001

Using International Standard Book Numbers as Uniform Resource Names

Status of this Memo

Copyright Notice

Abstract

This document discusses how International Standard Book Numbers (ISBN) can be supported within the URN (Uniform Resource Names) framework and the syntax for URNs defined in RFC 2141. Much of the discussion below is based on the ideas expressed in RFC 2288.

1. Introduction

As part of the validation process for the development of URNs, the IETF URN working group agreed that it is important to demonstrate that the current URN syntax proposal can accommodate existing identifiers from well established namespaces. One such infrastructure for assigning and managing names comes from the bibliographic community. Bibliographic identifiers function as names for objects that exist both in print and, increasingly, in electronic formats. RFC 2288 [Lynch, et al.] investigated the feasibility of using three identifiers (ISBN, ISSN and SICI) as URNs. This document will analyse the usage of ISBNs as URNs in more detail than RFC 2288.

A registration request for acquiring Namespace Identifier (NID) "ISBN" for ISBNs is included in chapter 5.

The document at hand is part of a global joint venture of the national libraries to foster identification of electronic documents in general and utilisation of URNs in particular. The document was written as a co-operative project between the Helsinki University Library and The International ISBN Agency.

We have used the URN Namespace Identifier "ISBN" for ISBNs in examples below.

2. Identification vs. Resolution

As a rule the ISBNs identify finite, manageably-sized objects, but these objects may still be large enough that resolution into a hierarchical system is appropriate.

The materials identified by an ISBN may exist only in printed or other physical form, not electronically. The best that a resolver will be able to offer in this case is bibliographic data from a national bibliography database, including information about where the physical resource is stored in the national library’s holdings.

3. International Standard Book Numbers

3.1 Overview

RFC 2288 [Lynch] describes the ISBN system in the following way:

> An International Standard Book Number (ISBN) identifies an edition of a monographic work. The ISBN is defined by the standard NISO/ANSI/ISO 2108:1992 [ISO1]
>
> Basically, an ISBN is a ten-digit number (actually, the last digit can be the letter "X" as well, as described below) which is divided into four variable length parts usually separated by hyphens when printed. The parts are as follows (in this order):
>
> * a group identifier which specifies a group of publishers, based on national, geographic or some other criteria,
>
> * the publisher identifier,
>
> * the title identifier,
>
> * and a modulus 11 check digit, using X instead of 10.

> The group and publisher number assignments are managed in such a way that the hyphens are not needed to parse the ISBN unambiguously into its constituent parts. However, the ISBN is normally transmitted and displayed with hyphens to make it easy for human beings to recognize these parts without having to make reference to or have knowledge of the number assignments for group and publisher identifiers.

Groups usually cover only one country, but occasionally a single group is used in several countries. For instance, group "3" is utilised in Germany, Austria and German-speaking parts of Switzerland. "976" is used in Caribbean community (Antigua, Bahamas, Barbados, Belize, Cayman Islands, Dominica, Grenada, Guyana, Jamaica, Montserrat, St. Kitts-Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago, Virgin Islands (Br))and "982" in South Pacific (Cook Islands, Fiji, Kiribati, Marshall Islands, Nauru, Niue, Solomon Islands, Tokelau, Tonga, Tuvalu; Vanuatu, Western Samoa). For each international group, the International ISBN Agency has assigned ranges of publisher identifiers to individual countries. These ranges are listed on the ISBN web site (http://www.isbn.spk-berlin.de/html/prefix.htm). The group identifiers are listed at http://www.isbn.spk-berlin.de/html/prefix/allpref.htm.

There are plans to extend the ISBN into 13 digits in order to make the system more suitable for identification of electronic monographs. So called Bookland ISBN will consist of a traditional ISBN preceded by the 978 or 979 EAN flag.

3.2 Encoding Considerations and Lexical Equivalence

RFC 2288 [Lynch] says that:

> Embedding ISBNs within the URN framework presents no particular encoding problems, since all of the characters that can appear in an ISBN are valid in the identifier segment of the URN. %-encoding, as described in [MOATS] is never needed.
>
> Example: URN:ISBN:0-395-36341-1
>
> For the ISBN namespace, some additional equivalence rules are appropriate. Prior to comparing two ISBN URNs for equivalence, it is appropriate to remove all hyphens, and to convert any occurrences of the letter X to upper case.

3.3 Resolution of ISBN-based URNs

The existing ISBN structure is suitable for URN resolution purposes. The group identifier can assist in the resolver discovery process. For instance, the group identifier "951" means Finland. In this case, the Finnish national bibliographic database will be able to resolve the URN either into bibliographic data or - if the resource is available in the Internet - to the document itself.

If a group identifier does not identify a single country but a language area, there are two means for locating the correct national bibliography. First, it is possible to define a cascade of URN resolution services - for instance, German national bibliography, Austrian national bibliography and Swiss national bibliography, in this order - into the DNS records describing the resolution service for ISBNs starting with "3". Second, the publisher identifier ranges assigned by the International ISBN Agency could be defined into the DNS records. This method is better than cascading, since the correct resolution service can be found immediately.

In some exceptional cases - notably in the US and in UK, where international companies do a significant portion of publishing - the information provided by the group identifier may not always be fully reliable. For instance, some monographs published in New York by international publishing companies may get an ISBN with the group identifier "3". This is technically appropriate when the headquarters or one of the offices of the publisher is located in Germany.

Information about such a book will not be available in the German national bibliography, but via the Library of Congress systems. Unfortunately, the appropriate national bibliography cannot be known to the resolver discovery service.

As a fall back mechanism a large union catalogue, such as WorldCat maintained by OCLC (http://www.oclc.org) could be used to complement the default services provided by national bibliographies.

The problem described above may well be less severe than it looks. Some international publishers (Springer, for example) give the whole production to the national library of their home country as legal deposit, no matter which country the book was published. Thus everything published by Springer in New York with group identifier "3" will be found from the German national bibliography. On the other hand, when these companies give their home base also as a place of publication, the "home" national library requires the legal deposit.

Due to the intelligent structure of ISBN, group identifier or even the publisher identifier can be used as a "hint". Technically, it is possible to incorporate into the common structure also URN resolution services maintained by publishers. For instance, "951-0" is the unique ISBN publisher identifier of the largest publisher in Finland, Sanoma-WSOY. If they launch their own URN resolution services, resolution requests for ISBNs starting with "951-0" will be directed to the publisher's server, and all other requests to the national bibliography.

3.4 Additional considerations

The basic guidelines for assigning ISBNs to electronic resources are the following:

* Format/means of delivery is irrelevant to the decision whether a product needs an ISBN or not. If the content meets the requirement, it gets an ISBN, no matter what the format of the delivery system.

* Each format of a digital publication should have a separate ISBN.

The definition of a new edition is normally based on one of the two criteria:

* A change in the kind of packaging involved: the hard cover edition, the paperback edition and the library-binding edition would each get a separate ISBN. The same applies to different formats of digital files.

* A change in the text, excluding packaging or minor changes such as correcting a spelling error. Again, this criterion applies regardless of whether the publication is in printed or in digital form.

Although these rules seem very clear, their interpretation may vary. As [Lynch] points out,

> The choice of whether to assign a new ISBN or to reuse an existing one when publishing a revised printing of an existing edition of a work or even a revised edition of a work is somewhat subjective. Practice varies from publisher to publisher (indeed, the distinction between a revised printing and a new edition is itself somewhat subjective). The use of ISBNs within the URN framework simply reflects these existing practices. Note that it is likely that an ISBN URN will often resolve to many instances of the work (many URLs).

Publishers have also in some occasions re-used the same ISBN for another book. This reasonably rare kind of human error does not threaten or undermine the value of the ISBN system as a whole. Neither do they pose a serious threat to the URN resolution service based on ISBNs. An error will only lead into the retrieval of two or more bibliographic records from a national bibliographic database. Based on the information in the records, a user can choose the correct record from the result set.

Most national bibliographies and especially the Books in Print correct ISBN mistakes. The systems then provide cross references ("incorrect ISBN -> correct ISBN").

Further details on the process of assigning ISBNs can be found in section 5 (Namespace registration) below.

4. Security Considerations

This document proposes means of encoding ISBNs within the URN framework. ISBN-based URN resolution service is depicted here only in a fairly generic level; thus questions of secure or authenticated resolution mechanisms are excluded. It does not deal with means of validating the integrity or authenticating the source or provenance of URNs that contain ISBNs. Issues regarding intellectual property rights associated with objects identified by the ISBNs are also beyond the scope of this document, as are questions about rights to the databases that might be used to construct resolvers.

5. Namespace registration

URN Namespace ID Registration for the International Standard Book Number (ISBN)

This registration describes how International Standard Book Numbers (ISBN) can be supported within the URN framework.

Namespace ID:

ISBN

This Namespace ID is the same as the internationally known acronym for the International Standard Book Number. Giving NID "ISBN" to any other identifier system would cause a lot of confusion.

Registration Information:

Version: 1
Date: 2001-01-25

Declared registrant of the namespace:

Name: Hartmut Walravens
E-mail: hartmut.walravens@sbb.spk-berlin.de
Affiliation: Director, The International ISBN Agency
Address: Staatsbibliothek zu Berlin - Preussischer Kulturbesitz -
D-10772 Berlin, Germany

Declaration of syntactic structure:

An ISBN is a ten-digit number (actually, the last digit can be the letter "X" as well, as described below) which is divided into four variable length parts usually separated by hyphens when printed. The parts are as follows (in this order):

* a group identifier which specifies a group of publishers, based on national, geographic or some other criteria,

* the publisher identifier,

* the title identifier,

* and a modulus 11 check digit, using X instead of 10.

Example:

URN:ISBN:0-395-36341-1

Relevant ancillary documentation:

The ISBN (International Standard Book Number) is a unique machine-readable identification number, which marks any edition of a book unambiguously. This number is defined in ISO Standard 2108. The number has been in use now for 30 years and has revolutionised the international book-trade. 154 countries are officially ISBN members, and more countries are joining the system.

The administration of the ISBN system is carried out on three levels:

International agency
Group agencies
Publisher levels

The International ISBN agency is located within the State Library Berlin. The main functions of the International ISBN Agency are:

* To promote, co-ordinate and supervise the world-wide use of the ISBN system.

* To approve the definition and structure of group agencies.

* To allocate group identifiers to group agencies.

* To advise on the establishment and functioning of group agencies.

* To advise group agencies on the allocation of international publisher identifiers.

* To publish the assigned group numbers and publishers prefixes in up-to-date form.

More information about ISBN usage can be found from the ISBN Users' Manual. 4th edition of this document is available at http://www.isbn.spk-berlin.de/html/userman.htm.

Identifier uniqueness considerations:

ISBN that has been assigned once should never be re-used. Nevertheless, publishers do occasionally re-use the same number. From the point of the URN resolution system proposed here, this will typically cause retrieval of two bibliographic records. A user can choose the correct publication using the data in the record, such as the author or title.

Incorrect ISBNs are routinely corrected in national bibliographies and Books in Print catalogue.

Identifier persistence considerations:

The ISBN accompanies a publication from its production onwards. It is persistent; ISBN once given - if correct - will never leave the publication.

Identifier assignment process:

Assignment of ISBNs is always controlled by ISBN group agencies, which are often national and quite frequently located in the national libraries. Publishers are usually given blocks of ISBNs, from which they pick identifiers for their newly published items.

As pointed out earlier, in spite of the common rules of how to use ISBNs, there is some variation between different publishers in ISBN assignment. In practice these differences are so small that they do not pose a threat to the usability of the ISBN system.

Identifier resolution process:

URNs based on ISBNs will be primarily resolved via the national bibliography databases. Since ISBN group agencies are as a rule located in national libraries, the national bibliography databases cover almost every publication which does have an ISBN.

If group identifier does not define a country but a language area there may be many countries using the same group identifier. In such cases, the International ISBN Agency has divided publisher identifiers into ranges assigned to each country within the group. The appropriate resolution service can be found by using the group identifier and publisher identifier information. Alternatively a cascade of national bibliographies can be defined.

Resolution carried out in national bibliography databases may be complemented by so called union catalogues, which contain huge amount of bibliographic data (up to 42 million records). This complementary service is only needed if the ISBN group identifier information is misleading. This is not common.

The International ISBN Agency maintains a list of publishers who have been assigned a publisher identifier within the ISBN system. The publisher identifier may be used to allow participation of resolution services maintained by publishers into the URN resolution system for ISBN.

Rules for Lexical Equivalence:

For the ISBN namespace, some additional equivalence rules are appropriate. Prior to comparing two ISBN URNs for equivalence, it is appropriate to remove all hyphens, and to convert any occurrences of the letter X to upper case.

Conformance with URN Syntax:

Embedding ISBNs within the URN framework presents no particular encoding problems, since all of the characters that can appear in an ISBN are valid in the identifier segment of the URN %-encoding, as described in [MOATS] is never needed.

Example: URN:ISBN:0-395-36341-1

Validation mechanism:

Validity of an ISBN string can be checked by modulus 11 check digit, included in the ISBN. X is used instead of 10.

Validity of ISBN assignments can be checked from the group agencies or directly from the publisher.

Scope:

Global.

6. References

[Daigle] Daigle, L., van Gulik, D., Iannella, R. and P. Faltstrom, "URN Namespace Definition Mechanisms", RFC 2611, June 1999.

[Lynch] Lynch, C., Preston, C. and R. Daniel, "Using Existing Bibliographic Identifiers as Uniform Resource Names", RFC 2288, February 1998.

[Moats] Moats, R., "URN Syntax", RFC 2141, May 1997.

7. Authors' Addresses

Juha Hakala
Helsinki University Library - The National Library of Finland
P.O. Box 26
FIN-00014 Helsinki University
FINLAND

EMail: juha.hakala@helsinki.fi

Hartmut Walravens
The International ISBN agency
Staatsbibliothek zu Berlin - Preussischer Kulturbesitz -
D-10772 Berlin
GERMANY

EMail: hartmut.walravens@sbb.spk-berlin.de

8. Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.