

Using National Bibliography Numbers as Uniform Resource Names

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2001). All Rights Reserved.

Abstract

This document discusses how national bibliography numbers (persistent and unique identifiers assigned by the national libraries) can be supported within the URN (Uniform Resource Names) framework and the syntax for URNs defined in RFC 2141. Much of the discussion is based on the ideas expressed in RFC 2288.

1. Introduction

As part of the validation process for the development of URNs the IETF 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] investigated the feasibility of using three identifiers (ISBN, ISSN and SICI) as URNs.

This document will analyse the usage of national bibliography numbers (NBNs) as URNs. The need to extend analysis to new identifier systems was briefly discussed in RFC 2288 as well, with the following summary: "The issues involved in supporting those additional identifiers are anticipated to be broadly similar to those involved in supporting ISBNs, ISSNs, and SICIs".

A registration request for acquiring a Namespace Identifier (NID) "NBN" for national bibliography numbers has been written by the National Library of Finland on the request of the Conference of Directors of National Libraries (CDNL) and the Conference of the European National Librarians (CENL). Chapter 5 contains a URN namespace registration request modeled according to the template in RFC 2611.

The document at hand is part of a global co-operation of the national libraries to foster identification of electronic documents in general and utilisation of URNs in particular. Some national libraries, including the national libraries of Finland, Norway and Sweden, are already assigning NBN-based URNs for electronic resources.

We have used the URN Namespace Identifier "NBN" for the national bibliographic numbers in examples below.

2. Identification vs. Resolution

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

The materials identified by a national bibliography number 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 a national library's holdings.

The URN Framework provides resolution services that may be used to describe any differences between the resource identified by a URN and the resource that would be returned as a result of resolving that URN. However, NBNs will be used for instance to identify resources in digital Web archives created by harvester robot applications. In this case, NBN will identify exactly the resource the user expects to see.

3. National bibliography numbers

3.1 Overview

National Bibliography Number (NBN) is a generic name referring to a group of identifier systems utilised by the national libraries and only by them for identification of deposited publications which lack an identifier, or to descriptive metadata (cataloging) that describes the resources. In many countries legal (or voluntary) deposit is being extended to electronic publications.

Each national library uses its own NBN strings independently of other national libraries; there is no global authority which controls them. For this reason NBNs are unique only on national level. When used as URNs, NBN strings must be augmented with a controlled prefix such as country code. These prefixes guarantee uniqueness of the NBN-based URNs on the global scale.

NBNs have traditionally been given to documents that do not have a publisher-assigned identifier, but are cataloged to the national bibliography. NBNs can be seen as a fall-back mechanism: if no other, better established identifier such as ISBN can be given, an NBN is assigned. In principle, NBN usage enables identification of any Internet document. Local policies may limit the NBN usage to a much smaller subset of documents.

Some national libraries (e.g., Finland, Norway, Sweden) have established Web-based URN generators, which enable authors and publishers to fetch NBN-based URNs for their network documents. At least national libraries of Sweden and Finland are harvesting and archiving domestic Web documents (and a number of other libraries plan to start this activity), and long-time preservation of these materials requires persistent and unique identification. NBNs can be and are in fact already used as internal identifiers in these Web archives.

Both syntax and scope of NBNs can be decided by each national library independently. Typically, an NBN consist of one or more letters and/or digits. This simple syntax makes NBNs infinitely extensible and very suitable for e.g., naming of the Web documents. For instance the application used by the national library of Finland for Web harvesting creates NBNs which are based on the MD5 checksum of the archived resource.

3.2 F-code

F-code is the NBN used by the National Library of Finland.

F-codes have been used since early 20th century to identify catalogue cards and later MARC records in the national bibliography. In 1998 the national library decided to enable the Finnish authors and publishers to assign F-codes to their Internet documents, if these documents do not qualify for other identifiers such as ISBN. F-codes, embedded into URNs, can be fetched from the URN generator (<http://www.lib.helsinki.fi/cgi-bin/urn.pl>) developed in co-operation between the national library of Finland and the Lund University library, NETLAB unit. Attached to the generator there is a user guide (<http://www.lib.helsinki.fi/meta/URN-opas.html>; only in Finnish), which tells the users how to use URNs.

F-codes are also used within the Web harvesting and archiving software (<http://www.csc.fi/sovellus/nedlib/>), which has been built for the Networked European Deposit Library (NEDLIB) project (see <http://www.kb.nl/nedlib>). NEDLIB harvester calculates MD5 checksum for each archived resource, and then builds an NBN-based URN from the checksum. The URN serves then as a unique identifier to the archived resource. Traditional identifiers can not be used for this purpose, since there may for instance be several variants of a book which (quite rightly so) all have the same ISBN. Moreover, identifiers embedded into a document do not necessarily belong to the document itself; thus the Web archiving application can not trust the identifiers embedded into the body of the document.

The F-code built by the URN generator consist of:

Prefix (for example fe)
Year (YYYY; for example 1999)
Number (for example 1055)

The generator also adds namespace identifier "NBN" and ISO 3166 country code. Thus a URN based on F-code would in this case be for instance `urn:nbn:fi-fe19991055`.

URNs created by the Web archiving application have similar overall structure, except that prefix (which may be defined by the operator) is fea and year is not used. An example: `urn:nbn:fi-fea-5c5875e6e49ae649cad63e5ee4f6c346`.

F-codes never need any special encoding when used as URNs, since they consist of alphanumeric codes only (0-9, a-z). This is often the case for other national libraries' NBN systems as well.

3.3 Encoding Considerations and Lexical Equivalence

Embedding NBNs within the URN framework usually presents no particular encoding problems, since all of the characters that can appear in commonly used NBN systems can be expressed in special encoding, as described in RFC 2141 [MOATS].

When an NBN is used as a URN, the namespace specific string will consist of three parts: prefix, consisting of either a two-letter ISO 3166 country code or other registered string, delimiting character which is either hyphen (-) or colon (:), and NBN string assigned by the national library. Delimiting characters are not lexically equivalent.

Hyphen is always used for separating the prefix and the NBN string.

Colon is used as the delimiting character if and only if a country code-based NBN namespace is split further in smaller sub-namespaces. If there are several national libraries in one country, these libraries can split their national namespace into smaller parts using this method.

A national library may also assign a trusted organisation(s) its own sub-namespace. For instance, the national library of Finland has given Statistics Finland (http://www.stat.fi/index_en.html) a sub-namespace "st" (e.g., urn:nbn:fi:st:). The Finnish Council of State (<http://www.vn.fi/vn/english/index.htm>) will use sub-namespace "vn" (e.g., urn:nbn:fi:vn).

Non-ISO 3166-prefixes, if used, must be registered on the global level. The Library of Congress will maintain the central register of reserved codes. This register will be available to the national libraries and other users in the Web.

Sub-namespace codes beneath a country-code-based namespace need to be registered on the national level by the national library which assigned the code. The national register must be available in the Web and should also be linked to the global register maintained by the Library of Congress.

Two-letter codes may not be used as non-ISO prefixes, since all such codes are reserved for existing and possible future ISO country codes. If there are several national libraries in one country who use the same prefix - for instance, a country code -, they need to agree on how to split the namespace between them.

Models:

URN:NBN:<ISO 3166 country code>-<assigned NBN string>

URN:NBN:<ISO 3166 country code>:<sub-namespace code>-<assigned NBN string>

URN:NBN:<non-ISO 3166 prefix>-<assigned NBN string>

Examples:

URN:NBN:fi-fe19981001 (A "real" URN assigned by the National Library of Finland).

3.4 Resolution of NBN-based URNs

The (usually) country code-based prefix part of the URN namespace specific string will provide a guide to where to find a resolution service, and the NBN register will identify the assigning agency. Once the NBN-based URN resolution is in global usage, the number of prefixes will slowly approach and may eventually exceed the number of national libraries.

If NBN assignment for a given country is limited to the national bibliography database, then all NBN-based URNs for that country will be resolved there. In one model these databases contain detailed resource descriptions including URLs, which will point both to the copy of the document in the Internet and to the copy in the national library's (legal) deposit collection. Due to the limitations in the usage of legal deposit documents it is possible that the deposited electronic materials can not be delivered in electronic form outside the premises of the national library.

If it is possible for the authors and publishers to retrieve NBNs to Web documents and there is no obligation to deposit thus identified documents to the national library, URN resolution service is not possible without a national Web index and archive, maintained by the national library or other organisation(s). A Web index/archive will also resolve machine-generated URNs to the archived Web documents.

3.5 Additional considerations

Guidelines adopted by each national library define when different versions of a work should be assigned the same or differing NBNs. These rules apply only if identifier assignment is done manually. If identifiers are allocated programmatically, the only criteria that can be used is that two documents which are identical on the bit level (have the same MD5 checksum) are deemed identical and should receive the same NBN. The likelihood of this happening to dissimilar documents is about 2^{-64} , according to the RFC 1321.

The rules governing the usage of NBNs are less strict than those specifying the usage of ISBN or other, better established identifiers. Since the NBNs have up to now been given only by the personnel (cataloguers) working in the national libraries, the identifier assignment has in practice been well co-ordinated.

A NBN-based URN will resolve to single instance of the work if identifier assignment has been automatic. Given the nature of NBNs it is also likely that different versions of the same work will receive different NBNs even if the identifier is given manually.

4. Security Considerations

This document proposes means of encoding several existing bibliographic identifiers within the URN framework. This document does not discuss resolution except at a very generic level; thus questions of secure or authenticated resolution mechanisms are out of scope. It does not address means of validating the integrity or authenticating the source or provenance of URNs that contain bibliographic identifiers. Issues regarding intellectual property

rights associated with objects identified by the various bibliographic identifiers 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 National Bibliography Number (NBN)

Namespace ID:

NBN

This Namespace ID has been in production use in demonstrator systems since summer 1998; thousands of URNs from this namespace have already been delivered in Finland, Sweden and Norway.

Registration Information:

Version: 3

Date: 2001-01-30

The first registration of the NID "NBN" was done via the URN WG in 1998. The second, slightly edited registration request was done in 1999.

Declared registrant of the namespace:

Name: Juha Hakala

E-mail: juha.hakala@helsinki.fi

Affiliation: Helsinki University Library - The National Library of Finland, Conference of European National Librarians (CENL) and Conference of Directors of National Libraries (CDNL)

Address: P.O.Box 26, 00014 Helsinki University, Finland

Both CENL and CDNL made decisions to foster the usage of URNs during 1998. The latter organisation has set up a working group for this purpose. One item in the common work plan is utilisation of national bibliography numbers as URNs for identification of grey literature published in the Internet. The NBN namespace will be available for free for all national libraries in the world.

Declaration of syntactic structure:

The namespace specific string will consist of three parts:

prefix, consisting of either a two-letter ISO 3166 country code or other registered string and sub-namespace codes,

delimiting characters (colon (:), or hyphen (-), and

NBN string assigned by the national library.

Colon is used as a delimiting character only within the prefix, between ISO 3166 country code and sub-namespace code, which splits the national namespace into smaller parts. This technique can be used when there are several national libraries, which all need their own namespaces, or when the national library allows trusted partners to set up their own sub-namespaces within the national NBN namespace.

Dividing non-ISO 3166-based namespaces further with sub-namespace codes is not allowed.

Hyphen is used as a delimiting character between the prefix and the NBN string. Within the NBN string, hyphen can be used for separating different sections of the code from one another.

Non-ISO prefixes used instead of the ISO country code must be registered. A global registry, maintained by the Library of Congress, will be created and made available via the Web. Contact information: nbn.register@loc.gov.us.

All two-letter codes are reserved for existing and possible future ISO country codes and may not be used as non-ISO prefixes.

Sub-namespace codes must be registered on the national level by the national library which assigned the code. The register must be available via the Web, and it should be accessible via the global registry set up by the Library of Congress.

Models:

URN:NBN:<ISO 3166 country code>-<assigned NBN string>

URN:NBN:<ISO 3166 country code:sub-namespace code>-<assigned NBN string>

URN:NBN:<non-ISO 3166 prefix>-<assigned NBN string>

Example:

A country code-based URN: URN:NBN:fi-fe19981001 (A URN assigned by the National Library of Finland).

Relevant ancillary documentation:

National Bibliography Number (NBN) is a generic name referring to a group of identifier systems used by the national libraries for identification of deposited publications which lack an identifier, or to descriptive metadata (cataloguing) that describes the resources. Each national library uses its own NBN system independently of other national libraries; there is no global authority which controls syntax of these identifier systems.

Each national library can decide freely which resources will receive NBNs. These identifiers have traditionally been assigned to documents that do not have a publisher-assigned identifier, but are nevertheless catalogued to the national bibliography. Typically identification of grey publications have largely been dependent on NBNs.

Some national libraries (Finland, Norway, Sweden) have established Web-based URN generators, which enable authors and publishers to fetch NBN-based URNs for their network documents.

Both syntax and scope of NBNs is decided by each national library independently. Typically, a NBN consist of one or more letters and a number.

Identifier uniqueness considerations:

NBN strings assigned by two national libraries may be identical. For this reason usage of a controlled prefix in the namespace specific string is obligatory in order to guarantee global uniqueness of NBN-based URNs.

In the national level, libraries utilise different policies for guaranteeing uniqueness. A national library may automate the delivery of NBN-based URNs. In this case, the NBNs are assigned sequentially by a program (URN generator).

Identifier persistence considerations:

Persistence of the NBNs as identifiers is guaranteed by the persistence of national libraries and information systems, such as national bibliographies, maintained by them. NBNs have been used for several centuries for printed materials. NBN-based identification of electronic documents is a recent practice, but it is likely to continue for a very long time.

Process of identifier assignment:

Assignment of NBN-based URNs is always controlled on national level by the national library / national libraries. The Conference of Directors of National Librarians (CDNL) has established in 1999 a task force, which will co-ordinate the URN usage in all national libraries.

National libraries may choose different strategies in assigning NBN-based URNs. One option is assignment by the library personnel only. This is done when the document is catalogued into the national bibliography. Thus in this case the national bibliography database will serve as the URN resolution service.

A national library may also set up a URN generator (generators), and allow publishers and authors to retrieve NBN-based URNs from there. In this case there is no guarantee that the identified resource will ever be catalogued into the national bibliography, and URN resolution is dependent on Web index/archive.

Process for identifier resolution:

URNs based on NBNs will be primarily resolved via the national bibliography databases. In one model these databases contain detailed resource descriptions including URLs, which will point both to the copy of the document in the Internet and to the copy in the national library's (legal) deposit collection. Due to the limitations in the usage of legal deposit documents it is possible that the deposited materials can not be delivered outside the premises of the national library.

For those documents not catalogued into the national bibliography database URN resolution may take place via national or international Web indexes and/or archives. Nordic national libraries have established in autumn 2000 a joint initiative called Nordic Web Archive (NWA), which aims at creating a national Web archive into all Nordic countries. Indexes to these archive systems will be able to act as URN resolution services of any document which a) is or has been available via the Web, and b) had an URN embedded into it.

Country code and additional sub-namespace information will provide a guide to where to find appropriate resolution services. For instance, if the country code is "fi", the primary resolution service is the national bibliography database. Secondary resolution service is the Web archive.

Generally, there will be one or more resolution services specified for each country, depending on the assignment policy and services of the national library. If NBN assignment is limited to the national bibliography database, then all NBN-based URNs for that country will be resolved there. If the authors and publishers have been allowed to retrieve NBNs to their Web resources, URN resolution services require a national Web archive. If other organisations have been allowed to assign NBNs, they may also set up their own URN resolution services.

Rules for Lexical Equivalence:

None in the global level. Any national library may provide its own rules, on the basis of its NBN syntax.

Conformance with URN Syntax:

All NBNs we know of are ASCII strings consisting of letters (a-z) and numbers (0-9). If NBN contains characters that are reserved in the URN syntax, this data must be presented in hex encoded form as defined in RFC 2141. A national library may limit the full scope of its NBN strings in URN usage in such a way that there are no reserved characters in the URN namespace specific strings.

Validation mechanism:

None specified on the global level. A national library may use NBNs, which contain a checksum and can therefore be validated, but this is for the time being not a common practice.

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. Author's Address

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

EMail: juha.hakala@helsinki.fi

8. Full Copyright Statement

Copyright (C) The Internet Society (2001). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

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