Using National Bibliography Numbers as Uniform Resource Names

Abstract

National Bibliography Numbers (NBNs) are used by national libraries and other organizations in order to identify resources in their collections. NBNs are usually applied to resources that are not catered for by established (standard) identifier systems such as International Standard Book Number (ISBN).

A Uniform Resource Name (URN) namespace for NBNs was established in 2001 in RFC 3188. Since then, a number of European national libraries have implemented URN:NBN-based systems.

This document replaces RFC 3188 and defines how NBNs can be supported within the updated URN framework. A revised namespace registration (version 4) compliant to RFC 8141 is included.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8458.
Table of Contents

1. Introduction ........................................... 3
2. Conventions Used in This Document ...................... 5
3. Fundamental Namespace and Community Considerations for NBN  5
   3.1. The URN:NBN Namespace .............................. 5
   3.2. Community Considerations for NBNs ................. 6
4. National Bibliography Number URNs ........................ 7
   4.1. Assignment .......................................... 7
   4.2. Syntax ................................................ 8
       4.2.1. Usage of r-component and q-component ........ 10
       4.2.2. Usage of f-component .......................... 10
   4.3. Encoding Considerations and Lexical Equivalence .... 10
   4.4. Resolution and Persistence of NBN-based URNs ....... 12
   4.5. Additional Considerations ........................... 13
5. URN Namespace ID (NID) Registration for the National
   Bibliography Number (NBN) .............................. 13
6. IANA Considerations ....................................... 15
7. Security Considerations ................................... 15
8. References ............................................... 15
   8.1. Normative References ............................... 15
   8.2. Informative References .............................. 16
Appendix A. Significant Changes from RFC 3188 ............... 18
Acknowledgements .......................................... 18
Contributors .............................................. 18
Author’s Address ........................................... 18
1. Introduction

One of the basic permanent Uniform Resource Identifier (URI) schemes (cf. [RFC3986] and [IANA-URI]) is Uniform Resource Name (URN). URNs were originally defined in RFC 2141 [RFC2141]. In 2017, a revision was adopted with new definitions and registration procedures [RFC8141]. Any traditional identifier, when used within the URN system, must have a namespace of its own that is registered with IANA [IANA-URN]. National Bibliography Number (NBN) is one such namespace, specified in 2001 in RFC 3188 [RFC3188].

This document describes the syntax and usage of NBN URNs and updates the registration of the associated URN namespace. This document additionally describes certain policy assumptions about how national libraries and their partner organizations partition, delegate, and manage the namespace. Violation of those assumptions could impact the utility of the NBN URN namespace.

URN:NBNs are in production use in several European countries including (in alphabetical order) Austria, Finland, Germany, Hungary, Italy, the Netherlands, Norway, Sweden, and Switzerland. The URN:NBN namespace is collectively managed by these national libraries. URN:NBNs have been applied to diverse content including Web archives, digitized materials, research data, and doctoral dissertations. They can be used by national libraries and organizations cooperating with them.

As a part of the initial development of the URN system in the late 1990s, the IETF URN Working Group agreed that it was important to demonstrate that the URN syntax can accommodate existing identifier systems. RFC 2288 [RFC2288] investigated the feasibility of using ISBN, ISSN, and SICI (Serial Item and Contribution Identifier) as URNs, with positive results; however, it did not formally register corresponding URN namespaces. (For further discussion of how these systems have evolved as URNs, see RFC 8254 [RFC8254].) This was in part due to the still-evolving process to formalize criteria for namespace definition documents and registration. The criteria were consolidated later in the IETF, first in RFC 2611 [RFC2611], then RFC 3406 [RFC3406], and now RFC 8141 [RFC8141].

URN namespaces have been registered for NBN, ISBN, and ISSN in RFCs 3188 [RFC3188], 3187 [RFC3187], and 3044 [RFC3044], respectively. ISBN and ISSN namespaces were made compliant with RFC 8141 [RFC8141] in 2017 by publishing revised ISSN [ISSN-namespace] and ISBN [ISBN-namespace] namespace registrations.
The term "National Bibliography Number" encompasses persistent local identifier systems that national libraries and their partner organizations use in addition to the more formally (and internationally) established identifiers. These partner organizations include universities and their libraries and other subsidiaries, other research institutions, plus governmental and public organizations. Some national libraries maintain a significant number of these liaison relationships; for instance, the German National Library had almost 400 by early 2018 [NBN-Resolving].

In practice, NBN differs from standard identifier systems such as ISBN and ISSN because it is not a single identifier system with standard-specified scope and syntax. Each NBN implementer creates its own system with its own syntax and assignment rules. Each user organization is also obliged to keep track of how NBNs are being used; however, within the generic framework set in this document, local NBN assignment policies may vary considerably.

Historically, NBNs have been applied in the national bibliographies to identify the resources catalogued into them. Prior to the emergence of bibliographic standard identifiers in the early 1970s, national libraries assigned NBNs to all catalogued publications.

Since the late 1990s, the NBN scope has been extended to cover a vast range of resources, both originally digital and digitized. Only a small subset of these resources is catalogued in the national bibliographies or other bibliographic databases. Digitized resources and their component parts (such as still images in books or journal articles) are examples of resources that may get NBNs.

It is possible to extend the scope of the NBN much further. The National Library of Finland is using them in the Finnish National Ontology Service Finto to identify corporate names (see <http://finto.fi/cn/en/>). Using NBNs to identify metadata elements provides a stable basis for creation of linked data.

Simple guidelines for using NBNs as URNs and the original namespace registration were published in RFC 3188 [RFC3188]. The RFC at hand replaces RFC 3188; sections discussing the methods by which URN:NBNs should be resolved have been updated, unused features have been eliminated, and the text is compliant with the stipulations of the revised URN specification [RFC8141].
2. Conventions Used in This Document

"NBN" refers to any National Bibliography Number identifier system used by the national libraries (or equivalent organizations) and other institutions, which use these identifiers with national libraries’ support and permission.

In this memo, "URN:NBN" is used as a shorthand for "NBN-based URN".

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Fundamental Namespace and Community Considerations for NBN

3.1. The URN:NBN Namespace

NBNs are widely used to identify both hand-held and digital resources in the collections of national libraries and other institutions that are responsible for preserving the cultural heritage of their constituents. Resources in these collections are usually preserved for a long time (i.e., for centuries). While the preferred methods for digital preservation may vary over time and depend on the content, the favorite one has been migration. Whenever necessary, a resource in an outdated file format is migrated into a more modern file format. To the extent possible, all old versions of the resource are also kept in order to alleviate the negative effects of partially successful migrations and the gradual loss of original look and feel that may accompany even fully successful migrations. When NBN is used to identify manifestations and there are many of them for a single work, local policy can require that each manifestation ought to have its own NBN.

NBNs are typically used to identify objects for which standard identifiers such as ISBN are not applicable. However, NBNs can be used for component resources even when the resource as a whole qualifies for a standard identifier. For instance, if a digitized book has an ISBN, JPEG image files of its pages might be assigned NBNs. These URN:NBNs can be used as persistent links to the pages.

The scope of standard identifier systems such as ISBN and ISSN is limited; they are applicable only to certain kinds of resources. One of the roles of the NBN is to fill in the gaps left by the standard identifiers. Collectively, these identifiers and NBNs cover all resources that national libraries and their partners need to include in their collections.
Section 4 below, and particularly Section 4.1, present a more detailed overview of the structure of the NBN namespace, related institutions, and the identifier assignment principles used.

3.2. Community Considerations for NBNs

National libraries are the key organizations providing persistent URN resolution services for resources identified with NBNs, independent of their form. As coordinators of NBN usage, national libraries have allowed other organizations, such as university libraries or governmental organizations, to assign NBNs to the resources these organizations preserve for the long term. In such case, the national library coordinates the use of NBNs at the national level. National libraries can also provide URN resolution services and technical services to other NBN users. These organizations are expected to either establish their own URN resolution services or use the technical infrastructure provided by the national library. URN:NBNs are expected to be resolvable and support one or more resolution services.

Although NBNs can be used to identify component resources, the NBN namespace does not specify a generic, intrinsic syntax for doing that. However, there are at least two different ways in which component resources can be taken into account within the NBN namespace.

The simplest and probably the most common approach is to assign a separate NBN for each component resource, such as a file containing a digitized page of a book, and make no provisions to make such NBNs discernible in a systematic way from others.

Second, if the stipulations of the URI generic syntax [RFC3986] and the Internet media type specification [RFC2046] are met, in accordance with the provisions in RFC 8141 [RFC8141], the URN f-component can be attached to URN:NBNs in order to indicate the desired location within the resource supplied by URN resolution.

From the library community point of view, it is important that the f-component is not a part of the Namespace-Specific String (NSS), and therefore f-component attachment does not mean that the relevant component part is identified. Moreover, the resolution process still retrieves the entire resource even if there is an f-component. The component part selection is applied by the resolution client (e.g., browser) to the resource returned by the resolution process. In other words, in this latter case the component parts are just logical and physical parts of the identified resource whereas in the former cases they are independently named entities.
Resources identified by NBNs are not always available in the Internet. If one is not, the URN:NBN can resolve to a surrogate such as a metadata record describing the identified resource.

Section 4 below, and particularly Section 4.4, presents a detailed overview of the application of the URN:NBN namespace as well as the principles of, and systems used for, the resolution of NBN-based URNs.

4. National Bibliography Number URNs

4.1. Assignment

National Bibliography Number (NBN) is a generic term referring to a group of identifier systems administered by national libraries and institutions authorized by them. The NBN assignment is typically performed by the organization hosting the resource. National libraries are committed to permanent preservation of their deposit collections.

Assignment of NBN-based URNs is controlled on a national level by the national library (or national libraries, if there is more than one). National guidelines can differ, but the identified resources themselves are usually persistent.

Different national URN:NBN assignment policies have resulted in varying levels of control of the assignment process. Manual URN:NBN assignment by the library personnel provides the tightest control, especially if the URN:NBNs cover only resources catalogued into the national bibliography. In most national libraries, the scope of URN:NBN is already much broader than this. Usage rules can vary within one country, from one URN:NBN sub-namespace to the next.

Each national library uses NBNs independently of other national libraries; apart from this document, there are no guidelines that specify or control NBN usage. As such, NBNs are unique only on the national level. When used as URNs, base NBN strings MUST be augmented with a controlled prefix, which is the particular nation’s ISO 3166-1 alpha-2 two-letter country code (referred to as "ISO country code" below) [ISO3166-1]. These prefixes guarantee uniqueness of the URN:NBNs at the global scale [ISO3166MA].

National libraries using URN:NBNs usually specify local assignment policies for themselves. Such policy can limit the URN:NBN usage to, e.g., the resources stored in the national library’s digital collections or databases. Although this specification does not
specify principles for URN:NBN assignment policies that can be applied. NBNs assigned to short-lived resources should not be made URN:NBNs unless such policy can be justified.

URN:NBN assignment policy can clarify, for instance, the local policy concerning identifier assignment to component parts of resources and can specify, with sufficient detail, the syntax of local component identifiers (if there is one as a discernible part of the NBNs). The policy can also cover any employed extensions to the default NBN scope.

NBNs as such are locally but not globally unique; two national libraries can assign the same NBN to different resources. A prefix, based on the ISO country code as described above, guarantees the global uniqueness of URN:NBNs. Once an NBN has been assigned to a resource, it MUST be persistent, and therefore URN:NBNs are persistent as well.

A URN:NBN, once it has been generated from a NBN, MUST NOT be reused for another resource.

Users of the URN:NBN namespace MUST ensure that they do not assign the same URN:NBN twice. Different policies can be applied to guarantee this. For instance, NBNs and corresponding URN:NBNs can be assigned sequentially by programs in order to avoid human mistakes. It is also possible to use printable representations of checksums such as SHA-1 [RFC6234] as NBNs.

4.2. Syntax

The Namespace-Specific String (NSS) will consist of three parts:

- a prefix consisting of an ISO 3166-1 alpha-2 country code and optional sub-namespace code(s) separated by a colon(s);
- a hyphen (-) as the delimiting character; and,
- an NBN string assigned by the national library or sub-delegated authority.
The following formal definition uses ABNF [RFC5234].

```
nbn-nss     = prefix "-" nbn-string
prefix      = iso-cc *( ":" subspc )
            ; The entire prefix is case insensitive.
iso-cc      = 2ALPHA
            ; Alpha-2 country code as assigned by part 1 of ISO 3166
            ; (identifies the national library to which the branch
            ; is delegated).
subspc      = 1*(ALPHA / DIGIT)
            ; As assigned by the respective national library.
nbn-string  = path-rootless
            ; The "path-rootless" rule is defined in RFC 3986.
            ; Syntax requirements specified in RFC 8141 MUST be
            ; taken into account.
```

A colon SHOULD be used within the prefix only as a delimiting character between the ISO 3166-1 country code and sub-namespace code(s), which splits the national namespace into smaller parts.

The structure (if any) of the nbn_string is determined by the authority for the prefix. Whereas the prefix is regarded as case insensitive, NBN strings can be case sensitive at the preference of the assigning authority; parsers therefore MUST treat these as case sensitive, and any case mapping needed to introduce case insensitivity is the responsibility of the relevant resolution system.

A hyphen SHOULD be used as the delimiting character between the prefix and the NBN string. Within the NBN string, a hyphen MAY be used for separating different sections of the identifier from one another.

All two-letter codes are reserved by the ISO 3166 Maintenance Agency for either existing or possible future ISO country codes (or for private use).

Sub-namespace identifiers MUST be registered on the national level by the national library that assigned the identifier. The list of such identifiers can be made publicly available via the Web.

Note that because case mapping for ASCII letters is completely reversible and does not lose information, the case used in case-insensitive matching is a local matter. Implementations can convert
to lower or upper case as they see fit; they only need to do it consistently.

4.2.1. Usage of r-component and q-component

URN:NBN resolvers do not currently support the use of either r-component or q-component.

Resolution services based on r-component can be implemented in the future when the r-component syntax and semantics have been specified.

4.2.2. Usage of f-component

If URN:NBN resolves to the identified resource and the media type of the resource supports f-component usage, it can be used to indicate a location within the identified resource. Persistence is achieved if the URN:NBN is assigned to one and only one version of a resource, such as a PDF/A version of a book.

The URN:NBN namespace does not impose any restrictions of its own on f-component usage.

4.3. Encoding Considerations and Lexical Equivalence

Expressing NBNs as URNs is usually straightforward, as normally only ASCII characters are used in NBN strings. If this is not the case, non-ASCII characters in NBNs MUST be translated into canonical form as specified in RFC 8141. If a national library uses NBNs that can contain percent-encoded characters higher than U+007F, the library needs to carefully define the canonical transformation from these NBNs into URNs, including normalization forms.

When an NBN is used as a URN, the NSS MUST consist of three parts:

- a prefix, structured as a primary prefix, which is a two-letter ISO 3166-1 country code of the library’s country, and zero or more secondary prefixes that are each indicated by a delimiting colon (:) and a sub-namespace identifier;
- a hyphen (-) as a delimiting character; and,
- the NBN string.

Different delimiting characters are not semantically equivalent.

The syntax and roles of the three parts listed above are described in Section 4.2.
If there are several national libraries in one country, these libraries MUST agree on how to divide the national namespace between themselves using this method before the URN:NBN assignment begins in any of these libraries.

A national library MAY also assign URN:NBN sub-namespaces to trusted organizations such as universities or government institutions. The sub-namespace MAY be further divided by the partner organization. All sub-namespace identifiers used within a country-code-based namespace MUST be registered on the national level by the national library that assigned the code. The national register of these codes SHOULD be made available online.

Being part of the prefix, sub-namespace identifier strings are case-insensitive. They MUST NOT contain any colons or hyphens.

Formally, two URN:NBNs are lexically equivalent if they are octet-by-octet equal after the following (conceptional) preprocessing:

1. convert all characters in the leading "urn:nbn:" token to a single case;
2. convert all characters in the prefix (country code and its optional sub-divisions) to a single case; and,
3. convert all characters embedded in any percent-encodings to a single case.

Models (indicated line break inserted for readability):

URN:NBN:<ISO 3166 alpha-2 country code>-<assigned NBN string>
URN:NBN:<ISO 3166 alpha-2 country code>:<sub-namespace code>-<assigned NBN string>

Examples:

URN:NBN:fi-fe201003181510
urn:nbn:ch:bel-9039
urn:nbn:se:uu:diva-3475
urn:nbn:hu-3006
4.4. Resolution and Persistence of NBN-based URNs

Eventually, URNs might be resolved with the help of a Global Resolver Discovery Service (GRDS), and URN:NBN syntax makes it possible to locate the relevant resolver. Since no GRDS system has been installed yet in the Internet, URN:NBNs are embedded in HTTP URIs in order to make them actionable in the present Internet. In these HTTP URIs, the authority part must point to the appropriate URN resolution service. For instance, in Finland, the address of the national URN resolver is <http://urn.fi>. Thus, the HTTP URI for the Finnish URN in the example above is <http://urn.fi/URN:NBN:fi-fe201003181510>.

The country-code-based prefix part of the URN:NBN namespace-specific string will provide a hint needed to find the correct resolution service for URN:NBNs from the GRDS when it is established.

There are three interrelated aspects of persistence that need to be discussed: persistence of the objects itself, persistence of the identifier, and persistence of the URN resolvers.

NBNs have traditionally been assigned to printed resources, which tend to be persistent. In contrast, digital resources require frequent migrations to guarantee accessibility. Although it is impossible to estimate how often migrations are needed, hardware and software upgrades take place frequently, and a lifetime exceeding 10-20 years can be considered as long.

However, it is a common practice to keep also the original and previously migrated versions of resources. Therefore, even outdated versions of resources can be available in digital archives, no matter how old or difficult to use they have become.

If all versions of a resource are kept, a user who requires authenticity can retrieve the original version of the resource, whereas a user to whom the ease of use is a priority is likely to be satisfied with the latest version. In order to enable the users to find the best match, a national library can link all manifestations of a resource to each other so as to make a user aware of them.

Thus, even if specific versions of digital resources are not normally persistent, persistent identifiers such as URN:NBNs support information architectures that enable persistent access to any version of the resource, including ones that can only be utilized by using digital archaeology tools such as custom-made applications to render the resource.
Persistence of URN resolvers themselves is mainly an organizational issue that is related to the persistence of organizations maintaining them. As URN:NBN resolution services will be supplied (primarily) by the national libraries, these services are likely to be long lived.

4.5. Additional Considerations

It is a good idea to apply URN:NBNs (or other persistent identifiers) to all resources that have been prioritized in the organization’s digital preservation plan.

Assignment of URN:NBNs to resources that are known to not be persistent should be considered carefully. URN:NBNs can, however, be applied to resources that have a low-level preservation priority and will not be migrated to more modern file formats or preserved via emulation.

If the identified version of a resource has disappeared, the resolution process can supply a surrogate if one exists. A surrogate can be, for instance, a more modern digital version of the original electronic resource.

5. URN Namespace ID (NID) Registration for the National Bibliography Number (NBN)

This URN namespace registration describes how National Bibliography Numbers (NBNs) can be supported within the URN framework; it uses the updated IANA template specified in RFC 8141.

Namespace Identifier: NBN

This namespace ID was formally assigned to the National Bibliography Number in October 2001, when the namespace was registered officially [RFC3188]. Utilization of URN:NBNs had started in demo systems already in 1998. Since 2001, tens of millions of URN:NBNs have been assigned. The number of users of the namespace has grown in two ways: new national libraries have started using NBNs, and many national libraries using the system have formed new liaisons.

Version: 4

Date: 2018-04-09
 Registrant:
   Name: Juha Hakala
   Affiliation: Senior Adviser, The National Library of Finland
   Email: juha.hakala@helsinki.fi
   Postal: P.O. Box 15, 00014 Helsinki University, Finland
   Web URL: http://www.nationallibrary.fi/

   The National Library of Finland registered the namespace on behalf of the Conference of the European National Librarians (CENL) and Conference of Directors of National Libraries (CDNL). The NBN namespace is available for free for the national libraries. They can allow other organizations to assign URN:NBNs and use the resolution services established by the library for free or for a fee. The fees, if collected, can be based on, e.g., the maintenance costs of the system.

Purpose: See Section 3 of RFC 8458

Syntax: See Section 4.2 of RFC 8458

Assignment: See Section 4.1 of RFC 8458

Security and Privacy: See Section 7 of RFC 8458

Interoperability:
   National libraries and their partners usually apply URN:NBNs if a standard identifier such as ISBN is not applicable for the resource to be identified. Some overlap with other URN namespaces is possible.

   URN:NBNs may contain characters which must be percent-encoded, but usually they consist of printable ASCII characters only.

Resolution: See Section 4.4 of RFC 8458

Documentation: RFC 8458

Revision Information:
   This version of the URN:NBN namespace registration has been updated to use the revised definition of URN syntax from RFC 8141, although usage of r-components is not specified yet. In addition, non-ISO 3166 (country code) based NBNs have been deleted due to lack of deployment. The entire NBN prefix is now specified to be case insensitive in accordance with established practice. This version also includes numerous clarifications based on actual usage of URN:NBNs.
6. IANA Considerations

IANA has updated the existing registration of the formal URN namespace, "NBN", using the template given above in Section 5.

7. Security Considerations

This document defines means of encoding NBNs as URNs. A URN resolution service for NBN-based URNs is depicted but only at a generic level; thus, questions of secure or authenticated resolution mechanisms and authentication of users are out of scope of this document.

Although no validation mechanisms are specified on the global level (beyond a routine check of those characters that require special encoding when employed in URIs), NBNs assigned by any given authority can have a well-specified and rich syntax (including, e.g., fixed length and checksum). In such cases, it is possible to validate the correctness of NBNs programmatically.

Issues regarding intellectual property rights associated with objects identified by the URN:NBNs are beyond the scope of this document, as are questions about rights to the databases that might be used to construct resolution services.

Beyond the generic security considerations laid out in the underlying documents listed in the Normative References, no specific security threats have been identified for NBN-based URNs.

8. References

8.1. Normative References


8.2. Informative References


Appendix A. Significant Changes from RFC 3188

Numerous clarifications have been made based on a decade of experience with RFC 3188.

NBNs that are not based on ISO 3166 (country codes) have been removed due to lack of usage.

In accordance with established practice, the whole NBN prefix is now declared case insensitive.

The document is based on the new URN syntax specification, RFC 8141.

Use of query components and fragment components with this namespace is now specified in accordance with RFC 8141.

Acknowledgements

Revision of RFC 3188 started during the project PersID [PERSID]. Later, the revision was included in the charter of the URNbis Working Group and worked on in that group in parallel with what became RFCs 8141 and 8254. The author wishes to thank his colleagues in the PersID project and the URNbis participants for their support and review comments.

Tommi Jauhiainen has provided feedback on an early draft version of this document. The author wishes to thank Tommi Jauhiainen, Bengt Neiss, and Lars Svensson for the comments they have provided to various draft versions of this document.

John Klensin provided significant editorial and advisory support for later draft versions of the document.

Contributors

This document would not have been possible without contributions by Alfred Hoenes.

Author’s Address

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

Email: juha.hakala@helsinki.fi