                  Using Existing Bibliographic Identifiers
                                    as
                         Uniform Resource Names

Status of this Memo

Copyright Notice

Abstract

   A system for Uniform Resource Names (URNs) must be capable of
   supporting identifiers from existing widely-used naming systems.
   This document discusses how three major bibliographic identifiers
   (the ISBN, ISSN and SICI) can be supported within the URN framework
   and the currently proposed syntax for URNs.

1. Introduction

   The ongoing work of several IETF working groups, most recently in the
   Uniform Resource Names working group, has culminated the development
   of a syntax for Uniform Resource Names (URNs).   The functional
   requirements and overall framework for Uniform Resource Names are
   specified in RFC 1737 [Sollins & Masinter] and the specification for
   the URN syntax is RFC 2141 [Moats].

   As part of the validation process for the development of URNs the
   IETF working group has 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. This memo demonstrates the feasibility of supporting three

representative bibliographic identifiers within the currently
proposed URN framework and syntax.

Note that this document does not purport to define the "official"
standard way of moving these bibliographic identifiers into URNs; it
merely demonstrates feasibility. It has not been developed in
consultation with these standards bodies and maintenance agencies
that oversee the existing bibliographic identifiers. Any actual
Internet standard for encoding these bibliographic identifiers as
URNs will need to be developed in consultation with the responsible
standards bodies and maintenance agencies.

In addition, there are several open questions with regard to the
management and registry of Namespace Identifiers (NIDs) for URNs.
For purposes of illustration, we have used the three NIDs "ISBN",
"ISSN" and "SICI" for the three corresponding bibliographic
identifiers discussed in this document. While we believe this to be
the most appropriate choice, it is not the only one. The NIDs could
be based on the standards body and standard number (e.g. "US-ANSI-
NISO-Z39.56-1997" rather than "SICI"). Alternatively, one could lump
all bibliographic identifiers into a single "BIBLIOGRAPHIC" name
space, and structure the namespace-specific string to specify which
identifier is being used. Any final resolution of this must wait for
the outcome of namespace management discussions in the working group
and the broader IETF community.

For the purposes of this document, we have selected three major
bibliographic identifiers (national and international) to fit within
the URN framework. These are the International Standard Book Number
(ISBN) [ISO1], the International Standard Serials Number (ISSN)
[NISO1,ISO2, ISO3], and the Serial Item and Contribution Identifier
(SICI) [NISO2]. An ISBN is used to identify a monograph (book). An
ISSN is used to identify serial publications (journals, newspapers)
as a whole.   A SICI augments the ISSN in order to identify
individual issues of serial publications, or components within those
issues (such as an individual article, or the table of contents of a
given issue). The ISBN and ISSN are defined in the United States by
standards issued by the National Information Standards Organization
(NISO) and also by parallel international standards issued under the
auspices of the International Organization for Standardization (ISO).
NISO is the ANSI-accredited standards body serving libraries,
publishers and information services. The SICI code is defined by a
NISO document in the United States and does not have a parallel
international standards document at present.

   Many other bibliographic identifiers are in common use (for example,
   CODEN, numbers assigned by major bibliographic utilities such as OCLC
   and RLG, national library numbers such as the Library of Congress
   Control Number) or are under development. While we do not discuss
   them in this document, many of these will also need to be supported
   within the URN framework as it moves to large scale implementation.
   The issues involved in supporting those additional identifiers are
   anticipated to be broadly similar to those involved in supporting
   ISBNs, ISSNs, and SICIs.

2. Identification vs. Resolution

   It is important to distinguish between the resource identified by a
   URN and the resources a URN resolver that can reasonably return when
   attempting to resolve an identifier. For example, the ISSN 0040-781X
   identifies the popular magazine "Time" -- all of it, every issue for
   from the start of publication to present. Resolving such an
   identifier should not result in the equivalent of hundreds of
   thousands of pages of text and photos being dumped to the user's
   machine. It is more reasonable for ISSNs to resolve to a
   navigational system, such as an HTML-based search form, so the user
   may select issues or articles of interest. ISBNs and SICIs, on the
   other hand, do identify finite, manageably-sized objects, but these
   objects may still be large enough that resolution to a hierarchical
   system is appropriate.

   In addition, the materials identified by an ISSN, ISBN or SICI may
   exist only in printed or other physical form, not electronically.
   The best that a resolver may be able to offer is information about
   where to get the physical resource, such as library holdings or a
   bookstore or publisher order form. 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.

3. International Standard Book Numbers

3.1 Overview

   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.

3.2 Encoding Considerations and Lexical Equivalence

   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 Additional considerations

   The ISBN standard and related community implementation guidelines
   define when different versions of a work should be assigned the same
   or differing ISBNs. In actuality, however, practice varies somewhat
   depending on publisher as to whether different ISBNs are assigned for
   paperbound vs. hardbound versions of the same work, electronic vs.
   printed versions of the same work, or versions of the same work
   distinguished in some other way (e.g., published for example in the
   US and in Europe). 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).

4. International Standard Serials Numbers

4.1 Overview

   International Standard Serials Numbers (ISSN) identify a work that is
   published on a continued basis in issues; they identify the entire
   (often open-ended, in the case of an actively published) work. ISSNs
   are defined by the international standards ISO 3297:1986 [ISO2] and
   ISO/DIS 3297 [ISO3] and within the United States by NISO Z39.9-1992
   [NISO1]. The ISSN International Centre is located in Paris and
   coordinates a network of regional centers. The National Serials Data
   Program within the Library of Congress is the US Center of this
   network.

   ISSNs have the form NNNN-NNNN where N is a digit, the last digit may
   be an upper case X as the result of the check character calculation.
   Unlike the ISBN the ISSN components do not have much structure;
   blocks of numbers are passed out to the regional centers and
   publishers.

4.2 Encoding Considerations and Lexical Equivalence

   Again, there is no problem representing ISSNs in the namespace-
   specific string of URNs since all characters valid in the ISSN are
   valid in the namespace-specific URN string, and %-encoding is never
   required.

   Example: URN:ISSN:1046-8188

   Supplementary comparison rules are also appropriate for the ISSN
   namespace. Just as for ISBNs, hyphens should be dropped prior to
   comparison and occurrences of 'x' normalized to uppercase.

4.3 Additional Considerations

   The ISSN standard and related community implementation guidelines
   specify when new ISSNs should be assigned vs. continuing to use an
   existing one. There are some publications where practice within the
   bibliographic community varies from institution to institution, such
   as annuals or annual conference proceedings. In some cases these are
   treated as serials and ISSNs are used, and in some cases they are
   treated as monographs and ISBNs are used. For example SIGMOD Record
   volume 24 number 2 June 1995 contains the Proceedings of the 1995 ACM
   SIGMOD International Conference on Management of Data. If you
   subscribe to the journal (ISSN 0163-5808) this is simply the June
   issue. On the other hand you may have acquired this volume as the
   conference proceedings (a monograph) and as such would use the ISBN
   0-89791-731-6 to identify the work. There are also varying practices

within the publishing community as to when new ISSNs are assigned due
to the change in the name of a periodical (e.g. Atlantic becomes
Atlantic Monthly); or when a periodical is published both in printed
and electronic versions (e.g. The New York Times). The use of ISSNs
in URNs will reflect these judgments and practices.

5. Serial Item and Contribution Identifiers

5.1 Overview

The standard for Serial Item and Contribution Identifiers (SICI)
codes, which has recently been extensively revised, is defined by
NISO/ANSI Z39.56-1997 [NISO2]. The maintenance agency for the SICI
code is the UnCover Corporation.

SICI codes can be used to identify an issue of a serial, or a
specific contribution (e.g., an article, or the table of contents)
within an issue of a serial. SICI codes are not assigned, they are
constructed based on information about the issue or issue component
in question.

The complete syntax for the SICI code will not be discussed here; see
NISO/ANSI Z39.56-1997 [NISO2] for details. However, an example and
brief review of the major components is needed to understand the
relationship with the ISSN and how this identifier differs from an
ISSN. An example of a SICI code is: 0015-
6914(19960101)157:1<62:KTSW>2.0.TX;2-F

The first nine characters are the ISSN identifying the serial title.
The second component, in parentheses, is the chronology information
giving the date the particular serial issue was published. In this
example that date was January 1, 1996. The third component, 157:1,
is enumeration information (volume, number) for the particular issue
of the serial. These three components comprise the "item segment" of
a SICI code. By augmenting the ISSN with the chronology and/or
enumeration information, specific issues of the serial can be
identified. The next segment, <62:KTSW>, identifies a particular
contribution within the issue. In this example we provide the
starting page number and a title code constructed from the initial
characters of the title. Identifiers assigned to a contribution can
be used in the contribution segment if page numbers are
inappropriate. The rest of the identifier is the control segment,
which includes a check character. Interested readers are encouraged
to consult the standard for an explanation of the fields in that
segment.

5.2 Encoding Considerations and Lexical Equivalence

   The character set for SICIs is intended to be email-transport-
   transparent, so it does not present major problems. However, all
   printable excluded and reserved characters from the URN syntax are
   valid in the SICI character set and must be %-encoded.

   Example of a SICI for an issue of a journal:

          URN:SICI:1046-8188(199501)13:1%3C%3E1.0.TX;2-F

   For an article contained within that issue:

          URN:SICI:1046-8188(199501)13:1%3C69:FTTHBI%3E2.0.TX;2-4

   Equivalence rules for SICIs are not appropriate for definition as
   part of the namespace and incorporation in areas such as cache
   management algorithms. It is best left to resolver systems which try
   to determine if two SICIs refer to the same content. Consequently,
   we do not propose any specific rules for equivalence testing through
   lexical manipulation.

5.3 Additional Considerations

   Since the serial is identified by an ISSN, some of the ambiguity
   currently found in the assignment of ISSNs carries over into SICI
   codes. In cases where an ISSN may refer to a serial that exists in
   multiple formats, the SICI contains a qualifier that specifies the
   format type (for example, print, microform, or electronic). SICI
   codes may be constructed from a variety of sources (the actual issue
   of the serial, a citation or a record from an abstracting service)
   and, as such are based on the principle of using all available
   information, so there may be multiple SICI codes representing the
   same article [NISO2, Appendix D]. For example, one code might be
   constructed with access to both chronology and enumeration (that is,
   date of issue and volume, issue and page number), another code might
   be constructed based only on enumeration information and without
   benefit of chronology. Systems that use SICI codes employ complex
   matching algorithms to try to match SICI codes constructed from
   incomplete information to SICI codes constructed with the benefit of
   all relevant information.

6. Security Considerations

   This document proposes means of encoding several existing
   bibliographic identifiers within the URN framework. This document
   does not discuss resolution; 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.

7. References

   [ISO1] NISO/ANSI/ISO 2108:1992 Information and documentation
          -- International standard book number (ISBN)
   [ISO2] ISO 3297:1986 Documentation -- International standard
          serial numbering (ISSN)
   [ISO3] ISO/DIS 3297 Information and documentation --
          International standard serial numbering (ISSN) (Revision of ISO
          3297:1986)
   [Moats] Moats, R., "URN Syntax", RFC 2141, May 1997.
   [NISO 1] NISO/ANSI Z39.9-1992 International standard serial
          numbering (ISSN)
   [NISO 2] NISO/ANSI Z39.56-1997 Serial Item and Contribution
          Identifier
   [Sollins & Masinter] Sollins, K., and L. Masinter, "Functional
          Requirements for Uniform Resource Names", RFC 1737, December
          1994.

8. Authors' Addresses

    Clifford Lynch
    Executive Director
    Coalition for Networked Information
    21 Dupont Circle
    Washington, DC 20036

    EMail: cliff@cni.org

    Cecilia Preston
    Preston & Lynch
    PO Box 8310
    Emeryville, CA 94662

    EMail: cecilia@well.com

    Ron Daniel Jr.
    Advanced Computing Lab, MS B287
    Los Alamos National Laboratory
    Los Alamos, NM, 87545

    EMail: rdaniel@acl.lanl.gov

9. Full Copyright Statement