

Network Working Group
Internet-Draft
Category: Informational
[draft-hakala-sici-01.txt](#)
Expires: 3 January 2003

Juha Hakala
Helsinki University Library
3 July 2002

Using Serial Item and Contribution Identifiers as Uniform Resource Names

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

To view the entire list of Internet-Draft Shadow Directories, see <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on 3 January 2003.

Abstract

This document discusses how Serial Item and Contribution Identifiers (SICIs; persistent and unique identifiers for serial issues and contributions such as articles) can be supported within the URN framework and the syntax for URNs defined in [RFC 2141](#) [[Moats](#)]. Much of the discussion below is based on the ideas expressed in [RFC 2288](#) [[Lynch](#)]. Chapter 5 contains a URN namespace registration request modelled according to the template in [RFC 2611](#) [Daigle et al.].

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 et. al.]

investigated the feasibility of using three identifiers (ISBN, ISSN and SICI) as URNs.

SICI is an American national standard defined by NISO/ANSI Z39.56-1996 [[NISO](#)]. The need to develop a new version of the standard is at present being investigated by NISO.

[RFC 2288](#) does not and it was not the aim of its authors to analyse how SICI-based URNs can actually be resolved. This text will specify one solution to this question. There may be other, complementary resolution services, in addition to the one described here.

Generally, the difficulty of designing a URN resolution service is dependent on two factors:

- * Is the identifier dumb, or does it provide a hint on where to find a resolution service?

- * How many potential resolution services are there?

ISBN (International Standard Book Number) is a good example of an intelligent identifier. Analysis of the ISBN will reveal not only the region where the ISBN has been assigned, but also the publisher of the book. Resolution of ISBN-based URNs can be decentralised to national bibliography databases, maintained by the national libraries. If the ISBN were a dumb identifier, this would be impossible.

International Standard Serial Number (ISSN) is a dumb identifier. It does not have a publisher identifier; serials published by a certain company get seemingly random ISSNs. Although ISSNs are allocated to regional agencies in blocks, which gives the system some "intelligence", a resolution service should not rely on these blocks - there are just too many of them, and their number is increasing all the time -, but use the global ISSN database. It contains a bibliographic description of every periodical that has received an ISSN; by June 2002 the database contained about one million bibliographic records. Thus, it is easy to resolve ISSN-based URNs even though the identifier itself does not help in localising the resolution service.

SICI is based on ISSN (see below for a description of its syntax). Like ISSN, it is therefore a dumb identifier. But there is not, and will never be, a global SICI database, which would contain bibliographic information about every serial issue and/or article published in the world. Most articles will not be catalogued at all, and the existing bibliographic information about articles is dispersed into a large number of databases maintained by publishers, libraries and other information intermediaries. Although it might be technically possible to merge records from these databases into a union catalogue, in practice such an enterprise is not politically possible.

As a "dumb" identifier with a large and ever growing number of potential

resolution services SICI poses interesting challenges to the design of the URN resolution process.

Generally, a combination of dumb identifier and multiple potential resolution services is a problem, since there is no simple way of finding out which resolution service is the correct one. A gateway service is needed for providing this valuable information. Below we propose that for SICI-based URNs, the global ISSN database could act as a link between the user and the resolution service.

The registration request for acquiring a Namespace Identifier (NID) "SICI" for Serial Item and Contribution Identifiers has been written by Helsinki University Library ù The National Library of Finland on behalf of the National Information Standards Organization (NISO). The request is included in chapter 5 of this text.

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. This work is co-ordinated by a working group established by the Conference of Directors of National Libraries (CDNL), and supported by the Conference of the European National Librarians (CENL) Working Group on Networking Standards.

We have used the URN Namespace Identifier "SICI" for the Serial Item and Contribution Identifiers in examples below.

2. Identification vs. Resolution

As a rule the SICIs identify finite, manageably-sized objects, but these objects may still be large enough so that resolution to a hierarchical system, such as all articles published in a serial issue, is appropriate.

The materials identified by a SICI may exist only in printed or other physical form, not electronically. The best that a resolver service will be able to offer in this case is bibliographic data from the database providing resolution services, including information about where the physical resource is stored in the owner institution's holdings.

3. Serial Item and Contribution Identifier

3.1 Overview

The Serial Item and Contribution Identifier (SICI) standard defines a variable length code that provides unique identification of serial items (e.g., issues) and the contributions (e.g., articles) contained in a serial title. SICI is specified in NISO/ANSI Z39.56-1996 [[NISO](#)]. Like other NISO standards, the SICI document is available for free in the Web.

SICI is based on ISO ISSN (International Standard Serial Number), but

augments it extensively. SICI is a combination of three segments, all of which are required:

Item segment; the data elements needed to describe the serial item such as serial issue (ISSN, Chronology, Enumeration)

Contribution segment, the data elements needed to identify contributions within an item (Location, Title Code)

Control segment, the data elements needed to record those administrative elements that determine the validity, version, and format of the SICI code representation.

[RFC 2288](#) provides the following example:

```
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.

SICI can be seen as a logical extension of the ISSN to the items and individual contributions that make up a serial's hierarchical structure. The current version of the SICI does have some limitations; it does not allow identification of subsections of an article such as paragraphs or diagrams. If deemed necessary, the functionality needed for article subsection identification could be added to the standard.

The current version of SICI (version 2, 1996) guarantees uniqueness in most situations; however, the standard does not always differentiate between multiple variant formats in which an electronic article may be published. For instance, variants of a digitised article published in PDF and HTML formats will receive the same SICI, provided that the ISSN is the same. There are plans to revise the standard; the new version may go further in allowing separation of different versions of an article from one another.

According to the rules of the ISSN centre, ISSN numbers can be applied retrospectively to old periodicals. If the original printed document has an ISSN, the same identifier is also valid for the digitised version. ISSN guidelines formulate this principle in the following way:

A reproduction is a copy of an item and intended to function as a substitute for that item. The reproduction may be in a different medium from the original but it is not a different edition in itself. The ISSN assigned to the original is valid for the reproduction, a new ISSN is not assigned to the reproduction.

ISSN numbers are assigned by regional agencies, which receive ISSN blocks from the ISSN International Centre. SICI usage is not dependent on such formal agencies; the aim is that once ISSN is known, SICI codes can be created by publishers, libraries, document delivery services or even by individual users, either manually or (preferably) by computer program.

Given the complexity of SICI codes, the recommended practice is to automate the SICI creation process. If an article is structured enough, all elements of SICI can be extracted from the document. A tool capable of this has been built by the E.U. project DIEPER; this tool, of course, only works properly if the document is structured in the way the DIEPER project recommends. Another, less challenging option is a SICI generator, which builds syntactically correct SICIs, including the check character, if the basic ingredients are typed in manually.

3.2 Encoding Considerations and Lexical Equivalence

[RFC 2288](#) contains the following simple and yet sufficient analysis of SICI encoding:

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.

3.3 Resolution of SICI-based URNs

Since ISSN is a dumb code, SICI does not contain any explicit hint on where to find the URN resolution service or services. However, an efficient and global resolution service can be accomplished by using the ISSN register as a way station. In June 2002, the ISSN register contained more than one million bibliographic records describing serials, including many thousands of electronic journals. There are several other databases, which contain hundreds of thousands of serial records, but the ISSN register has the best global coverage since the ISSN network covers more than 70 countries.

The first step in resolving a SICI-based URN is a query to the ISSN register. The SICI resolution service in the ISSN register will parse the SICI code in order to extract the ISSN from it.

ISSN will then be used as a search key for retrieving the bibliographic record of the serial from the ISSN register.

Currently the ISSN register already contains thousands of records describing electronic journals. These records contain the URL of the serial's home page.

This URL is appropriate for resolving the URN based on the ISSN of the periodical. The mechanism for resolving such URNs via the ISSN register has been specified in the [RFC 3044](#) [Rozenfeld]. The ISSN International Centre has built in 2001 a demonstration URN resolution service for ISSN-based URNs into their present information system, which will be replaced by a new integrated library system in 2003.

In order to resolve SICI-based URNs, a new data element has to be added into the bibliographic records in the ISSN register, or elsewhere in this database. This data element would contain the network address (URL) of the system (abstracting and indexing service), which holds the article required and/or bibliographic information about it. It must also be possible to specify volumes and if necessary issues which are included in the system. For example, one A & I service might contain volumes 1-50, while there is another system holding volumes 25-60. The data element should be repeatable, since the same article may be available from multiple sources. For instance, the publisher, Library of Congress (<http://www.loc.gov/>), JSTOR (<http://www.jstor.org/>) and a number of host services such as EBSCO (<http://www.ebsco.com/home/>) may all have a copy of the same resource.

Encoding the holdings data may be non-trivial, since this data is often highly volatile. So, although there are quite a few stable systems such as JSTOR, which can easily be encoded into the ISSN register, full coverage of A & I system will be unlikely.

The SICI resolution service built into the ISSN register will check if

database address information is available in the bibliographic record of the serial. As the next step, it can make sure that the volume and/or issue needed is available via the target service. If this is the case, the application will make the query, receive the result - article or bibliographic information about it - and pass it on to the user.

The functionality described above was implemented in co-operation between the ISSN International Centre and the E.U. project DIEPER (<http://gdz.sub.uni-goettingen.de/dieper/>). The SICI resolution service is an extension of the service built for resolving ISSN-based URNs. By March 2001 a demonstrator service via which several of the databases maintained by the project partners could be accessed was released for internal use within the project. The ISSN IC and project partners have been willing to maintain the service also after the formal end of the project.

Discussions about adding the new data element into bibliographic records in the ISSN register are as of this writing under way.

Please note that the discussion herein applies to SICIs assigned to serial contributions. Since serial items (issues) have seldom been described or digitised as such, a search by serial item SICI will in practice be expanded into retrieval of all contributions (articles) within the serial item (issue) in question.

If a resolution service for the resource at hand does not exist, or the user is not authorised to utilise it, he/she may get the bibliographic description of the serial from the ISSN register.

[3.4 Additional considerations](#)

Electronic journals have rapidly become very popular in scientific publishing. The main reasons for this are the emergence of viable business models (e.g. licensing) and the birth of a reliable and efficient delivery mechanism (the Web).

New content is being added via two different channels. A significant number of scientific journals are published as electronic versions alongside a printed version. On the other hand, old printed volumes are digitised and made available in electronic form. Digitisation is done by development projects such as DIEPER, established services such as JSTOR, or publishers - for instance Elsevier is digitising retrospectively all volumes of the journals the company has published.

Reliable linking of articles to references and bibliographic data about the articles is an important issue. URLs are as of this writing the most common means used for linking, but their reliability is low; average lifetime for a URL is estimated to be two years.

A more reliable linking mechanism than URLs is urgently needed. Many scientific publishers are already using Digital Object Identifiers (DOI)

for their materials. DOI resolution service is based on Handle system, which is "a comprehensive system for assigning, managing, and resolving persistent identifiers, known as "handles," for digital objects and other resources on the Internet" (see <http://www.handle.net/introduction.html>). Handles can be used as Uniform Resource Names(URNs).

URN is both an identifier and a non-commercial and technically advanced resolution service. Due to the co-operation of the ISSN International Centre the URN resolution service for articles outlined in this Internet standard is global, and can accommodate an unlimited number of article services located anywhere in the world.

For instance, in order to establish URN-based links to articles digitised in JSTOR service, a number of steps are necessary. First, each article must be identified by SICI, and these SICIs must be indexed in the JSTOR database. Second, the bibliographic records of JSTOR journals in the ISSN register must all be enriched with a link to the JSTOR search interface and volume/issue information. For instance, the bibliographic record describing the journal "Ecology" must contain the information that volumes 1-77 (1920-1996) are available via JSTOR. This information may be quite volatile, and maintenance of the ISSN register must therefore be frequent and efficient.

Apart from the modification of the data, some programming work is also needed. Due to the work done in the DIEPER project, the ISSN register already has the functionality needed for resolving SICI-based URNs. Adding the required functionality into the JSTOR database may or may not be difficult depending on the system architecture. One of the design aims of the URN system was to make building of resolution services easy and it seems that in this respect the designers were successful; in the DIEPER project some partners were able to implement the required functionality quite easily.

Since the Web browsers do not yet support URN resolution, the final step in enabling resolution of URN-based SICIs in DIEPER was the installation of the browser plug-in developed by the ISSN International Centre. In the future this step may not be required.

For various reasons, one article may be available in several locations. Every article copy may have a different set of users who are allowed access to it. For instance, a copy acquired by a national library via legal deposit may only be available within the library premises.

Making the links context sensitive ð to provide only those links that "work" for a user - is a challenge. OpenURL framework [Van de Sompel] provides a means for sensitive linking, and may be used to complement URN resolution service (filtering of those A & I -services which are not available to the user). As of this writing OpenURL is rapidly gaining popularity, and there are already a few integrated library systems

(MetaLib, Voyager) which support it. The future library system of the ISSN register may support OpenURL usage; this would be useful when the same resource (article) is available from several sources, which have different user population.

In their present form the URN resolution services provided via the ISSN register suit best those services, which are available in public domain, and are reasonably stable. Numerous digitisation projects such as DIEPER are currently making printed articles available in the Web in digital form.

An additional benefit of coding the needed location and volume information into the ISSN register would be that this database then could also serve as a global registry of serial digitisation efforts. Such a register is acutely needed to avoid duplicate work.

Since the number of SICI resolution services will eventually be high, the capacity of the server on which the ISSN register runs and its network connections may become a bottleneck, especially if the articles were delivered to the users via the ISSN server. Setting up mirror sites would in this case be the most efficient means for load control and balancing. Technically the setting up of mirror sites is not difficult. The ISSN register contains approximately a million bibliographic records, and is therefore not a very large database.

4. Security Considerations

This document proposes means of encoding and using Serial Item and Contribution Identifiers within the URN framework. This document does not discuss resolution except at a generic level; thus questions of secure or authenticated resolution mechanisms in the ISSN register or in actual resolution services are out of scope. This text does not address means of validating the integrity or authenticating the source or provenance of URNs that contain SICIs. 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 Serial Item and Contribution Identifier (SICI)

Namespace ID:

SICI

SICI is a well-established acronym for Serial Item and Contribution Identifiers; giving this NID for any other system would cause a lot of

confusion.

This namespace ID has already been used in SICI-based URNs in the E.U. project DIEPER.

Registration Information:

Version: 1

Date: 2002-07-03

Declared registrant of the namespace:

Name: Patricia Harris

E-mail: pharris@niso.org

Affiliation: National Information Standards Organisation

Address: 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814

Declaration of syntactic structure:

Each SICI contains three segments:

Item segment; the data elements needed to describe the serial item such as serial issue (ISSN, Chronology, Enumeration)

Contribution segment, the data elements needed to identify contributions within an item (Location, Title Code)

Control segment, the data elements needed to record those administrative elements that determine the validity, version, and format of the SICI code representation.

Example:

0015-6914(19960101)157:1<62:KTSW>2.0.TX;2-F

SICI codes can be generated and parsed by computer programs.

Relevant ancillary documentation:

SICI is an American national standard defined by NISO/ANSI Z39.56-1996 [NISO2]. A new version of the standard is currently under development.

Identifier uniqueness considerations:

SICI codes will almost always be unique. Since SICI is based on ISSN, articles from different journals will definitely never get the same SICI. Since enumeration and chronology information must also be given, articles and other contributions published in different volumes and issues will also never get the same SICI.

SICIs may not be unique if and only if:

If two or more contributions are published on the same page(s) and if they have similar enough titles (the first letter of each word is the same).

If in a single issue of an electronic journal (which lacks page numbers) there are two or more contributions with titles similar enough.

If there are several technical variants of an electronic serial contribution (multiple formats, multiple resolutions) the current version of SICI will not make any difference between these variants. In this case the intellectual content will usually be the same, but layout will differ from one version to another.

In the future the SICI standard may be enhanced in order to diminish the risk of non-unique SICIs.

Identifier persistence considerations:

Once assigned, SICI will never change. The same SICI will not be used again for other serial items and contributions.

Process of identifier assignment:

There will not be a national, regional or international agency governing the SICI assignment process. Publishers, libraries or other information intermediaries will create SICIs when needed. The most important prerequisite is that the journal must have an ISSN.

Although SICI assignment is decentralised, the national ISSN agencies and the ISSN International Centre may support publishers and other interested parties in SICI implementation.

SICI can - and should - be built via automated means. If the source document such as article is sufficiently structured, SICI can be generated without human involvement. Another option is a semi-automated process, in which a human user types in the relevant data elements and the application then builds the code.

Process for identifier resolution:

Resolution will take place in two steps as defined in chapter 3.3. First the ISSN register is used for finding the location of the resolution service(s) for the serial and volume at hand. Using the linking information stored in the serial's bibliographic record, the correct resolution service is contacted, and the requested resource is delivered to the user.

Rules for Lexical Equivalence:

We do not propose any specific rules for equivalence testing through lexical manipulation.

Conformance with URN Syntax:

According to the [RFC 2288](#):

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

Validation mechanism:

Validity of a SICI string can be checked by modulus 37 check digit.

Scope:

Global.

6. References

[Daigle et al.]: Daigle, L., van Gulik, D., Iannella, R. & Faltstrom, P.: URN Namespace Definition Mechanisms, [RFC2611](#), June 1999.

[Lynch] Lynch, C., Using Existing Bibliographic Identifiers as Uniform Resource Names, [RFC 2288](#), February 1998

[Moats] Moats, R., URN Syntax, [RFC 2141](#), May 1997.

[NISO] NISO/ANSI Z39.56-1996 Serial Item and Contribution Identifier. Electronic resource, available at <http://www.techstreet.com/cgi-bin/pdf/free/152629/z39-56.pdf>

[Rozenfeld] Rozenfeld, S., Using The ISSN (International Serial Standard Number) as URN (Uniform Resource Names) within an ISSN-URN Namespace, [RFC 3044](#), January 2001.

[Van de Sompel] Van de Sompel, Herbert & Beit-Arie, Oren: Open Linking in the Scholarly Information Environment Using the OpenURL Framework. D-

Lib Magazine, March 2001. Electronic resource, available at
<http://www.dlib.org/dlib/march01/vandesompe1/03vandesompe1.html>

7. Authors' Address

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

E-mail: juha.hakala@helsinki.fi

8. Full Copyright Statement

Copyright (C) The Internet Society (2002). 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.