

Network Working Group
Internet-Draft
Category: Informational
draft-ietf-hakala-istc-00.txt
Expires: 1 January 2003

Juha Hakala
Helsinki University Library
1 July 2002

Using International Standard Text Work Codes 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 1 January 2003.

Abstract

This document discusses how International Standard Text Work Codes (ISTCs; persistent and unique identifiers for textual works) can be supported within the URN framework and the syntax for URNs defined in RFC 2141 [Moats]. Analysis is in part based on the ideas expressed in RFC 2288 [Lynch], which analysed the use of ISSN, ISBN and SICI as URNs. 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 ISTC) as URNs.

As a result of a recent proliferation of manifestations of works (various printed and electronic versions of books, for instance) ISO has decided to develop a set of identifiers for works. These standards include International Standard Audiovisual Work Code (ISAN), International Standard Musical Work Code (ISWC) and International Standard Text Work Code (ISTC) [ISO].

These standards identify works (such as Brave New World by Aldous Huxley) and their manifestations (such as a translation of Brave new world into Finnish). Manifestations, like the first edition of the Brave new world by Chatto & Windus, London 1932, will never receive an ISTC but - it being a novel - an ISBN. ISTC and ISTC metadata will be efficient tools for bringing together all related works and expressions - like all translations of Brave new world - and all manifestations any work or expression may have.

ISTC is an emerging ISO standard which will reach the status of a Draft International Standard by Summer 2002. As of this writing it seems quite likely that the standard will be approved after the 6 months voting period in early 2003. Major changes to the syntax or to the maintenance organisation of the standard are very unlikely.

RFC 2288 does not - and it was not the aim of its authors - to analyse how ISTC-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.

Like ISBN, ISTC will be an intelligent identifier (see below for a description of its syntax). On the other hand, it will be similar to the ISSN system in that there will be a global ISTC database, containing every ISTC assigned in the world, and related metadata. Since ISTCs can and will be given to textual works retrospectively, this database, maintained by the ISTC Registration Authority, will relatively soon become very large.

However, at least some ISTC Regional Agencies, which will take care of ISTC assignment in their own regions (mainly geographical, but they may also be subject-driven) will send their data in batch mode to the ISTC register. Therefore there is a need to complement the ISTC resolution done in the global ISTC database with regional resolution services. The resulting system is a two-level cascade, where the bibliographic data related to the ISTC will be available either from the global database or from a database maintained by the Regional Agency, which assigned the ISTC. A Regional agency may be for instance a national library, which has generated work-related metadata and ISTCs from a traditional, manifestation-centered national bibliography.

The registration request for acquiring a Namespace Identifier (NID) "ISTC" for International Standard Text Work Codes has been written by Helsinki University Library - The National Library of Finland on behalf of the International Standardisation Organisation (ISO). 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 "ISTC" for the International Standard Text Work Codes in examples below.

2. Identification vs. Resolution

The ISTCs identify works, that is, abstract entities, which are embodied as physical manifestations. ISTC resolution service will only deliver a bibliographic record related to the work or expression. In the bibliographic record there may be links to other ISTC records describing related works and expressions, or to manifestations of the work.

The manifestations of textual works identified by ISTCs may be printed or electronic. In the latter case, a user may be able to retrieve all manifestations related to the work.

3. International Standard Text Work Code

3.1 Overview

The ISO International Standard Text Work Code (ISTC) standard defines a 16 byte hexadecimal code that provides unique identification of textual works. ISTC is as of this writing specified in the committee draft 21047, revised in 15 May 2002. In this CD, comments given to the first committee draft have been taken into account, and the ISTC Working Group decided to publish the text as a Draft International Standard. Changes to the syntax or management of the ISTC at this stage are highly unlikely.

ISTC consists of four segments, all of which are required:

- registration agency element;
- year element;
- work element;
- check digit.

ISO CD 21047 provides the following example:

ISTC 0A9 2002 12B4A105 7

When an ISTC is displayed in written form the letters ISTC shall precede it. The segments should be separated by hyphen or space.

Registration agency element shall consist of three hexadecimal digits. The code (in the above example, 0A9) represents the Registration agency which assigned the ISTC.

The year element (in the example, 2002) shall consist of the four digits representing the year in which the ISTC was allocated.

The work element shall consist of eight hexadecimal digits. The work element shall be assigned by an ISTC Registration agency appointed by the Registration authority for ISO 21047.

The check digit shall be calculated on a MOD 16-3 system defined in accordance with ISO 7064.

ISTC Registration agencies must provide metadata for each work they have identified. This metadata will be collected into the global ISTC register maintained by the ISTC Registration authority. The data may be updated on-line or in batch mode. Duplicates are removed from the database with the help of a duplicate check algorithm.

According to the ISO CD 21047, ISTCs can be applied retrospectively to old works. In such case, work metadata will be usually generated from existing manifestation level metadata. Some projects have already analysed the feasibility of this process with satisfactory results.

ISTC numbers are assigned by Registration agencies, which receive their agency element codes from the Registration authority. The system allows for 4096 such codes at any time; the codes may be re-used over time since agencies can be identified with the combination of the agency element and year. However, 4096 registration agency elements will be sufficient for quite a long time (the ISSN system has about 70 regional agencies, the ISBN system about 160).

Given the relative complexity of ISTC codes and the very large number of textual works, which need identification, the recommended practice is to automate the ISTC creation process. In any Registration agency the agency element will never change, and changes in the year element are easy to track. Work element can be used rather freely, as long as the same identifier is never assigned twice. Since calculation of the check digit can also be easily automated, ISTC assignment can without difficulty be made a fully automatic process.

3.2 Encoding Considerations and Lexical Equivalence

Since ISTC consists of hexadecimal characters, there are no needs for special encoding. However, the string ISTC preceding the identifier and any spaces separating the ISTC elements should be replaced by hyphens when an ISTC is used as URN.

In order to determine if two ISTCs are lexically equivalent it is necessary to remove all spaces and hyphens from the ISTC string.

3.3 Resolution of ISTC-based URNs

An efficient and global resolution service for ISTCs can be accomplished by using the global ISTC register. This database will, according to the current plans of the proposed Registration authority, go into production in January 2003. From this system, the ISTC data may be copied to one or several systems used for public access.

An ISTC can be used as a search key for retrieving the bibliographic record of the work from the databases containing ISTC data. This record may contain ISTCs pointing to other works or other identifiers such as ISBNs, DOIs or SICIs identifying manifestations (books or articles) of the textual work.

With the help of the registration agency element and the year code it is possible to locate the ISTC register (for instance, a traditional national bibliography database enriched with work metadata) of the Registration agency, which assigned the ISTC. Expanding the resolution of the ISTC-based URNs into these databases will bring two additional benefits. First, since the global ISTC register is maintained in batch mode it (and databases dependent on it) may not contain the newest ISTCs assigned by the registration authorities. Second, access to the systems containing global ISTC data may be for fee only, while the regional agencies may allow free access to their local ISTC registers.

Typical users of the system will be authors and publishers seeking information about (published or non-published) works, librarians wishing to copy catalogue metadata related to a given work, and patrons who wish to track all manifestations of a work or expression related to it.

3.4 Additional considerations

Since the number of ISTC resolution services will eventually be high (theoretical maximum 4096 + 1 "live" systems), encoding all services into the URN Resolution Discovery Service, and maintaining this data, may become a bottleneck.

The ISTC system may become very large, as it is intended to cover all textual works, including novels, short stories and articles. Such a system may eventually become extremely popular. It is important that there will be multiple databases containing all or at least the most of the ISTC metadata in existence.

4. Security Considerations

This document proposes means of encoding and using International Standard Text Work Codes 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 ISTC registers are out of scope. This text does not address means of validating the integrity or authenticating the source or provenance of URNs that contain ISTCs. Issues regarding intellectual property rights associated with bibliographic data related to the ISTC or other work 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 International Standard Text Work Code (ISTC)

Namespace ID:

ISTC

ISTC will become an established acronym for International Standard Text Work Codes; giving this NID for any other system would cause a lot of confusion.

Registration Information:

Version: 1

Date: 2002-07-01

Declared registrant of the namespace:

Name: International ISTC Agency / Albert Simmonds

E-mail: simmonda@oclc.org

Affiliation: OCLC Online Computer Library Center, Inc.

Address: OCLC, 6565 Frantz Road, Dublin, OH 43017-3395,
USA

Declaration of syntactic structure:

Each ISTC contains four segments:

ISTC consists of four segments, all of which are required:

- registration agency element;
- year element;

- work element;
- check digit.

When an ISTC is displayed in written form the letters ISTC shall precede it. The segments should be separated by hyphen or space.

Registration agency element shall consist of three hexadecimal digits. The code (in the above example, 0A9) represents the Registration agency which assigned the ISTC.

The year element (in the example, 2002) shall consist of the four digits representing the year in which the ISTC was allocated.

The work element shall consist of eight hexadecimal digits. The work element shall be assigned by an ISTC Registration agency appointed by the Registration authority for ISO 21047.

The check digit shall be calculated on a MOD 16-3 system defined in accordance with ISO 7064.

Example:

0A9-2002-12B4A105-7

ISTC codes can be generated and parsed by computer programs.

Relevant ancillary documentation:

ISTC is an emerging ISO standard defined by ISO CD 21047 (revised 2002-05-15). Draft International Standard version of ISTC will be published during summer 2002, and it is expected that ISTC will be approved as ISO standard in early 2003, after the DIS 6 months comment period. No major changes to the syntax of the ISTC or its maintenance organisation are likely.

Identifier uniqueness considerations:

ISTC codes will always be unique.

Two or more different ISTCs may identify the same work if multiple registration agencies deal with the same resources, or if a single agency deals with the same work twice. The duplicate control algorithm in the ISTC Registration authority is intended to remove duplicates arriving from the agencies, and any agency should have sufficient control mechanism in place to avoid duplicate registration of works.

Identifier persistence considerations:

Once assigned, ISTC will never change. The same ISTC will not be used again for another textual work.

Process of identifier assignment:

ISTCs will be assigned by the Registration agencies. Typically an author or his/her agent or a publisher will apply for an ISTC. It is also possible to generate ISTCs retrospectively for existing manifestations (published books and articles). This process has to be controlled well in order to avoid duplicate registration of works. One possibility is to generate work data in national bibliographic databases, and to limit the generation of work records to domestic works only.

The Registration authority will govern the ISTC assignment process in the global level. The global ISTC Registry will enable duplicate control of the identified works.

ISTC can - and should - be built via automated means.

Process for identifier resolution:

Resolution will take place as defined in chapter 3.3. The first step is to check the ISTC register or another database containing all of ISTC metadata, or the most of it. If there is no match, it is possible to use the Registration agency element (and eventually the year element) as a hint for finding the Registration agency, which has assigned the ISTC, and the resolution service maintained by it.

ISTCs will always resolve into the work metadata. Manifestations of the work (such as electronic versions of a book) may or may not be linked to the ISTC metadata. ISTC metadata may also contain links to related works and expressions.

Rules for Lexical Equivalence:

Spaces and hyphens in the ISTC string are lexically equivalent. String "ISTC" in the beginning of the string must be neglected in the comparison.

Conformance with URN Syntax:

ISTC consists of hexadecimal digits and it is therefore compliant to the requirements to the URN syntax as defined in [Moats].

Validation mechanism:

Validity of an ISTC string can be checked by modulus 16-3 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.

[ISO] Information and documentation - International Standard Text Code (ISTC). ISO/CD 21047. May 2002.

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

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

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.