

Draft Standard for Trial Use

Book Item and Component Identifier

Review Period: August 1, 2000 - January 31, 2002

A Draft Standard for Trial Use
Developed by the
National Information Standards Organization

Published by the National Information Standards Organization
Bethesda, Maryland



NISO Press, Bethesda, Maryland, U.S.A.

About NISO Standards

NISO Standards are developed by the Standards Committees of the National Information Standards Organization. The development process is a strenuous one that includes a rigorous peer review of proposed standards open to each NISO Voting Member and any other interested party. Final approval of the standard involves verification by the American National Standards Institute that its requirements for due process, consensus, and other approval criteria have been met by NISO. Once verified and approved, NISO Standards also become American National Standards.

The use of an ANSI/NISO Standard is voluntary. That is, the existence of this NISO Standard does not preclude anyone, whether or not that person has adopted the NISO Standard, from manufacturing, marketing, purchasing, or using products, processes, or procedures that do not conform to the NISO Standard. However, the use of the standards (those developed by NISO as well as other standards-developing organizations) has proven to be in the best interests of any industry wishing to increase its effectiveness and efficiency in the areas of product development, manufacturing, and marketing and, therefore, such use is encouraged by ANSI, NISO, and all other standards-developing organizations.

About NISO Draft Standards For Trial Use (DSFTU)

NISO releases proposed standards as Draft Standards for Trial Use (DSFTU) when there is a need for field experience.

Publication of this Draft Standard for Trial Use has been approved by NISO. Distribution of this draft Standard shall not continue beyond eighteen months from the date of publication. At the end of this time, this draft Standard will be revised as necessary and balloted, or withdrawn. Comments on its use or suggestions for revision should be directed to NISO. Questions on the status of this document should be directed to NISO.

This DSFTU has not been balloted by the NISO members and is not a consensus document. It is released for review and trial implementation for the period:

August 1, 2000 - January 31, 2002.

Comments, questions, and reports on trial implementations should be directed to:

National Information Standards Organization
4733 Bethesda Avenue, Suite 300
Bethesda, MD 20814-5248
Telephone: 301-654-2512 Fax: 301-654-1721
email: nisohq@niso.org
web page: www.niso.org

Copyright ©2000 by the National Information Standards Organization

This Draft Standard for Trial Use may be reproduced for noncommercial purposes to facilitate the review process. Address all inquiries to NISO.

Contents

Foreword	V
1. INTRODUCTION	1
1.1 PURPOSE	1
1.2 SCOPE	1
1.3 PRINCIPLES AND GUIDELINES	1
1.3.1 Use of ISBN	1
1.3.2 Self-derivability	1
1.3.3 Use of all available data and redundancy	2
1.3.4 Null data elements	2
1.4 STRUCTURAL MODEL	2
2. REFERENCED STANDARDS	3
3. DEFINITIONS	3
4. CHARACTER SET	5
5. PUNCTUATION	6
6. SEGMENTS AND DATA ELEMENTS	6
6.1 ITEM SEGMENT	7
6.1.1 ISBN	8
6.1.2 CHRONOLOGY	9
6.1.3 ITEM ENUMERATION	10
6.2 COMPONENT SEGMENT	11
6.2.1 Enumeration	11
6.2.2 Title Code	13
6.3 CONTROL SEGMENT	14
6.3.1 Code Structure Identifier (CSI)	14
6.3.2 Component type identifier (CTI)	15
6.3.3 Medium/Format Identifier (MFI)	17
6.3.4 Standard Version Number	18
6.3.5 Check Character	18
Appendix A ISBN Information	19
From the LARC Fact Sheet Number 28	19
ISBN Numbers	19
Appendix B Related Standards and References	20

Appendix C
Calculation of Modulus 37 Check Character 22

Appendix D
Usage Scenarios 23

Abstracting and Indexing Services 23
Document Delivery Services 23
Coursepack Production 23
Library Reserve Rooms 23
Rights Management Agencies 23
DOI System 23

Appendix F
% (Hex) Encoding 24

Encoding Requirements 24
Encoding Characters 24
Mandatory/Recommended Encoding 24

Figures and Tables

Figure 1: Block diagram of the BICI structural model 2
Table 1: Characters allowed in BICI 5
Table 2: Legal character usage by data element 5
Table 3: Use of Punctuation in BICI 6
Table 4: Relationships of Code Structure Identifiers (CSI) to other BICI segments 7
Table 5: Chronology codes 10
Table 6: CSI types 15
Table 7: CTI types 15
Table 8: MFI codes 17
Table C-1: Check Character Values 22
Table C-2: Example of Check Character Calculation 22
Table F-1 – Mandatory Encoding 25
Table F-2 – Recommended Encoding 25

Foreword

(This foreword is not part of the NISO Draft Standard for Trial Use, Book Item and Component Identifier. It is included for information only.)

This Book Item and Component Identifier (BICI) draft standard for trial use defines a variable length code that will provide unique identification of book items (e.g., where a single ISBN has been assigned to a multi-volume set) and to the component parts of a book (e.g., chapters, sections, illustrations, tables, etc.). It is intended primarily for use by those involved in the management of book components. While the BICI is applicable to both automated parsing and human-readable environments, it does not prescribe any specific machine-scannable symbology, nor does it prescribe a specific machine-readable interchange format for electronic transmission of the coded data.

As a draft standard for trial use, it defines the requirements for constructing a compact BICI code that has a unique value for each identifiable component of a book. Every effort has been made to specify requirements that can be applied to a vast majority of books, both in print and electronic form, and can be used in the many different applications to which the code is appropriate.

Background

The development of this draft standard began as an outgrowth of *ANSI/NISO Z39.56-1996, Serial Item and Contribution Identifier (SICI)*. A positive ballot vote for that standard resulted in inquiries from groups that wished to identify the contributions within non-serial items. Increased interest in item identification generally due to copyright and intellectual property rights management had raised awareness that in order to manage rights, the items had to be identified uniquely. These items could be of varying levels of granularity. An international meeting sponsored by IMPRIMATUR¹ and the Association of American Publishers called for the need for standard identifiers. In addition, the ISO Biblid standard² was withdrawn, leaving a void for any identifier standard that addressed non-serial items.

The work of the committee was initially based on several documents: BIBLID, SICI, and A Standard Identifier for Book Items and Contributions, a report prepared for Book Industry Communication (UK) and the British National Bibliography Research Fund by David Martin³ (June 1997). In addition, the committee was kept informed of the developments of international standards for work codes,⁴ audiovisual works,⁵ and music.⁶ The committee also took into account the work of INDECS⁷ and other emerging standards and specifications for electronic works.⁸ Committee members agreed it was important

¹ IMPRIMATOR "Issues: Identifying and Establishing Rights" <http://www.imprimator.alcs.uk/imprimatur/IRM/IMUSE.htm>

² ISO 9115 – 1987. "Documentation – Bibliographic identification (biblid) of contributions in serials and books.

³ ISBN: 1873671229. BLR&I Report number 82 and BNBRF Report 87.

⁴ ISO/CD 15707-1 International Standard Work Code (ISWC)

⁵ ISO/CD 15706 International Standard Audio Visual Number (ISAN)

⁶ ISO 10957:1993 International Standard Music Number (ISMN)

⁷ <http://www.indecs.org>

⁸ For example the DOI specification at <http://www.doi.org> and the Open e-Book work <http://www.openebook.org>

Book Item and Component Identifier

to build on existing standards, particularly the SICI, for those who were developing applications that would be taking advantage of identifier standards.

An initial set of discussions revolved around the issue of granularity. The standard as developed takes a conservative approach as it was designed to handle the traditional book structures (chapters, sections, preface, index, etc.) and typical citations (used most often in document delivery, academic coursepack and reserve activities) since those structures also appear in emerging electronic books.

In addition to the relationship to the SICI and BIBLID as discussed above, the BICI is one of a wide variety of identifier schemes currently in use or under-development. The two other identifier schemes that are most relevant here are the Document Object Identifier (DOI)⁹ and the Uniform Resource Name (URN).¹⁰

A more recent development, the Open Electronic Book (OEB) specification, provides another opportunity in which the BICI can be used to identify the component parts. The current OEB specification provides for identifiers at a level of granularity addressed by this standard.¹¹

A draft of this standard was posted on the NISO web-site for comments from the community from January 5, 2000 – April 17, 2000. Those comments are incorporated in this document.

Goals

The committee had several goals to guide the development of this standard:

- to limit the scope of the standard to a code for unique identification a book, and their component parts,
- to cover a broad range of books; for example, scholarly, trade, and popular, regardless of physical form,
- to maintain consistency with and build on the established standard of the ISBN (International Standard Book Number) already being used successfully to identify books, and
- to create a brief code consistent with unique identification.

This draft standard does not address gray literature, technical reports, or other documents. In the development of the standard, the committee determined that to completely address all types of documents, the standard would become too unwieldy. Several committee drafts did try to incorporate a variety of documents, but the final draft focuses on those publications that are most likely to take advantage of such an identifier. Building on the ISBN will provide a standard that addresses a variety of book publications.

⁹ DOI Handbook. version 0.3 July 2000. http://www.doi.org/handbook_2000/index.html

NISO Z39.84-2000 Syntax for the Digital Object Identifier

¹⁰ Lynch, C., C. Preston and R. Daniel, Jr. Using existing bibliographic identifiers as URNs (1998) RFC2288.

Moats, R. URN syntax (1997) RFC 2141 .

¹¹ <http://www.openebook.org>

After due consideration, the committee determined that this standard cannot provide for component identification when the item or component part has no enumeration, such as dictionary entries that are not numbered, multi-volume works in which the volume label is a text string. It is expected these works would not be used in the types of applications most likely intended to take advantage of the BICI. If turned into electronic works then these works would more likely follow a logical structure that could easily be identified.

The draft standard strives for uniqueness, but while the BICI speaks of a “unique” value for each unique component, the reader should understand that theoretically two components could have identical BICI values. The syntax and data elements specified in this standard attempt to minimize the occurrence of this situation, but duplicate values may occur. The committee believes this standard provides an acceptable balance between the conflicting design goals of uniqueness and compactness.

Although a unique and compact code was a goal, the committee has provided a certain amount of redundancy (e.g., ISBN and date, enumeration, title code and location) so that informed judgments can be made about the degree of similarity of two BICIs derived from different sources. The standard uses defined data elements and punctuation to allow easy visual parsing and machine recognition of the coding structure. There is no length restriction for a BICI.

The BICI is divided into three segments and, following the principle that the control segment governs the processing of the string, it uses a method similar to the SICI for specifying the coding structure. This data element is referred to as the Code Structure Identifier (CSI). The CSI is a numeric code at a specified position in the BICI string that allows both humans and computer systems to determine the type of coding. This standard defines three CSIs: CSI-1 specifies an item, CSI-2 denotes a component part identified through its logical structure, title code and through its physical structure (location); and CSI-3 provides a structure for private identifiers such as, but not limited to, sequentially assigned DOIs. This explicit identification of coding structure also provides for future extensibility.

Focus on the ISBN

The committee wrestled with its decision to limit this code to works assigned an ISBN, but determined that a workable code could only be developed by an extension of this commonly accepted standard. The ISBN is used to define the item. When creating codes for multi-volume works that have ISBN for the set, as well as the individual volumes use the ISBN for the volume. The committee also acknowledges that improper application of ISBN rules may result in incorrect BICIs, but that is something this standard cannot control. In addition, while there is no specific process in place for the assignment of retrospective ISBN, the committee is confident that those items most likely to be using BICIs will have ISBN. Discussion is currently underway in the ISBN community to change the ISBN to a thirteen-digit code to bring it into congruence with Bookland EAN.¹² Users of the BICI standard should be aware and monitor these developments.

¹² EAN International – the international numbering agency <http://www.ean.be>

Matching and Uniqueness

Extensive consideration was given by the committee to requirements for rights management, document supply, and interfile linking applications. It was recognized that the BICI must be derivable from the item in hand as well as any citations that might describe a component part of a book. Ideally, the best available source (the item itself) should be used to construct the enumeration and title code, and all available information should be included so the BICI is definitive and complete.

The construction of BICI codes from citations that appear in bibliographies does present problems. Some citation formats omit the information necessary to provide full chronology and enumeration, but the committee still recommended that both the logical and physical elements are mandatory if available.

The matching of BICI codes in some applications may require algorithms that are more complex than simple string comparisons. Since some data elements are mandatory if available when comparing BICI strings some data elements within one or both of the BICIs being matched may be omitted. This would be the case; for example, in document ordering applications where a patron-supplied code (derived from a citation, for example) that omits enumeration, could be matched against a BICI derived from either the item or a full-text database.

Integration with Other Standards

As with all NISO standards, references are given as ANSI, rather than ISO unless no ANSI standard exists.

Significant changes

Although, the BICI is an outgrowth of work done on the SICI there are a few differences in the structure, besides those discussed above.

Book Components

The committee reviewed a variety of items to locate examples to use within the draft standard. The draft that went out for community review used a code based on books with logical structure. The hierarchical structure of a work into its component parts – chapters, sections, charts, tables, etc. – are time-tested structures that are appearing in electronic works as well as the printed books with which the general user is most familiar. In that community reviewed version, the committee also provided a way to use location as a substitute for unknown logical structure. In investigations of citations, bibliographies, etc., it was often the case that the logical structure of a work was not represented but rather the location of that structure (e.g., pages 15-30 represented chapter 2).

After the community review, the structure of the component segment was changed to include both the logical and physical structure in the CSI-2. Both elements are mandatory if available. Therefore, CSI-3 is no longer needed to identify the use of location, and it is currently used to represent the proprietary number scheme that had been CSI-4 in the earlier drafts.

Electronic books as currently being described are being produced with logical structures and the committee expects that their component parts can be easily identified using the BICI.

Character set issues

Ideally, we would have liked to use Unicode¹³ in this identifier in order to facilitate broader international use. However, as we examined the issues involved in making this transition, it became clear that this was much more complex than it first seemed, and raised numerous interoperability and definitional problems. We have thus left the standard in 7-bit ASCII.

There are really only two sections of the standard that are sensitive to character set — the Enumeration and the Title Code. While it would be simple enough to simply say that these fields can contain arbitrary Unicode characters (with the exception of certain specific reserved characters) this does not get us far. It does raise questions about the contexts though which BICIs that contain Unicode can be transported with integrity. The use of Unicode at this time also introduces a number of issues for implementors such as Unicode canonicalization. Essentially, there are multiple ways to represent some characters in Unicode. So either the algorithm for forming a BICI has to specify a canonicalization, or else all systems using BICIs will need to canonicalize them before doing any additional comparisons or computations to test if two BICI codes refer to the same work. It is probably more efficient and less problematic to do this at the time that the BICI is generated, but this would introduce additional complexity into this standard that the committee felt ill prepared to undertake. The Unicode consortium¹⁴ and the Worldwide Web Consortium¹⁵ are currently struggling with these issues, and perhaps a future revision of this standard using Unicode might simply point at that work when it is completed.

More problematically, however, is the fact that the BICI (and the SICI) depend on language, not simply character sets. The title code extraction algorithms are designed for languages that use the Latin-1 character set.¹⁶ They rely on assumptions about words and word boundaries, punctuation, capitalization, etc., such as English, Spanish, French, or German; it is not at all clear that they are applicable to a broader range of languages that do not provide similar word breaks in the written language. We believe that the majority of the languages where we have confidence that the title code algorithm should produce reasonable results are actually handled acceptably by the current ASCII character set. The committee simply does not have the expertise to even begin to assess the title code algorithm for East Asian languages for example, other than to recognize that the current work is probably inadequate. Truly internationalizing of the identifier will require much more than just permitting the Unicode character set to be employed in the Title Code. We will need language-specific rules for extracting these title codes.

¹³ The Unicode Standard, version 3.0. Addison-Wesley, 2000. ISBN: 0-201-61633-5

¹⁴ <http://www.unicode.org>

¹⁵ <http://www.w3c.org>

¹⁶ ISO 8859-1:1987 (E) Latin alphabet no. 1.

ANSI/NISO Z39.47-1993. Extended Latin Alphabet Coded Character Set for Bibliographic use.

Book Item and Component Identifier

Similarly, it is not obvious what the use of the Unicode character set in the enumeration will accomplish. It raises ambiguities about whether page numbers, for example, should be translated into Arabic numbering symbols or whether they should be left in the vernacular.

These issues, combined with the limited expertise of the committee in the relevant areas, led us to the conclusion that full multilingual extension (which would require Unicode) represented problems we did not have the expertise to address. However, that simply permitting Unicode without addressing these questions would accomplish little except for adding more complexity and ambiguity.

Notational difference BICI / SICI

The use of the asterisk (*) instead of the hash sign (#) in the shelf-check code. In a report David Martin prepared for BIC and the BNBf dated 2 June 1997, in which David writes: "The SICI standard specifies the hash sign as the check character corresponding to the value 36. On some international keyboards, and in some national character set conventions, the ASCII code for the hash sign is, or may be, used for other purposes. It is recommended that it should not therefore be used as the BICI check character. Instead, the asterisk is suggested. Since this character is widely used as a mathematical operator in programming languages, it can be expected to be universally available."

Parenthesis (()) replace the angle brackets (<>) in the control segment. Many reviewers felt that the % encoding the angle bracket (<>) imposed too much overhead in the development of a BICI, and as such, the parenthesis should be used. URI and HTTP require the & encoding of restricted characters used in the string a list of this characters is available in Appendix F.

Other Items of Note

A maintenance agency will be selected as adoption of the standard matures. The role of the maintenance agency will include registration of the codes for the Code Structure Identifier (CSI), the Component type identifier (CTI), and the Medium/Format Identifier.

The standard does not prescribe any physical location for presenting the BICI on printed publications, nor does it prescribe any machine-scannable symbology.

This proposed national standard is being released as a Draft Standard for Trial Use for the period August 1, 2000 - January 31, 2002. At the end of this review period, the standard will be revised as necessary and balloted, or withdrawn. Comments on its use or suggestions for revision should be sent to the National Information Standards Organization, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814; email: nisohq@niso.org; telephone: 301-654-2512. Questions on the status of this document or the specifications described should be directed to NISO.

This DSFTU has not been balloted by the NISO members and is not a consensus document. It is released for review and trial implementation. At the time this proposed standard was released, NISO had the following Voting Members:

NISO Voting Members
3M

Jerry Karel
Carolyn Egeberg (Alt)

Academic Press

Anthony Ross
Bradford Terry (Alt)

ALSi

Martin Sach
Tony O'Brien (Alt)

American Association of Law Libraries

Robert L. Oakley
Mary Alice Baish (Alt)

American Chemical Society

Robert S. Tannehill, Jr.

American Library Association

Paul J. Weiss

American Medical Association

James A. Skowrenski
Vanessa Hayden (Alt)

American Society for Information Science

Mark H. Needleman

American Society of Indexers

Charlotte Skuster
Marie Kascus (Alt)

American Theological Library Association

Myron Chace

ARMA International

Diane Carlisle

Art Libraries Society of North America

David L. Austin

Association for Information and Image Management

Betsy A. Fanning

Association of Information and Dissemination Centers

Bruce H. Kiesel

Association of Jewish Libraries

Pearl Berger
Elizabeth Vernon (Alt)

Association of Research Libraries

Duane E. Webster
Julia Blixrud (Alt)

Baker & Taylor

Robert H. Doran

Book Industry Communication

Brian Green

Broadcast Music Inc.

Edward Oshanani
Robert Barone (Alt)

Cambridge Scientific Abstracts

Matthew Dunie
Anthea Gotto (Alt)

CARL Corporation

Ward Shaw

College Center for Library Automation

J. Richard Madaus
Ann Armbrister (Alt)

Committee on Institutional Cooperation

Thomas Peters

Congressional Information Service, Inc.

Robert Lester

Data Research Associates, Inc.

Michael J. Mellinger
Mark H. Needleman (Alt)

EBSCO Information Services

Melanie Watts

Elsevier Science Inc.

John Mancía

Endeavor Information Systems, Inc.

Verne Coppi
Cindy Miller (Alt)

Ex Libris

James Steenbergen
Carl Grant (Alt)

Follett Corp.

D. Jeffrey Blumenthal
Don Rose (Alt)

The Gale Group

Katherine Gruber
Justine Carson (Alt)

Gaylor Information Systems

James English
William Schickling (Alt)

GCA Research Institute

Jane Harnad

Geac Computers

Andrew Oates

Indiana Cooperative Library Services Authority

Millard Johnson
Dea Szatkowski (Alt)

Innovative Interfaces, Inc.

Gerald M. Kline
Sandy Westall (Alt)

Institute for Scientific Information

Helen Atkins
Richard Newman (Alt)

The International DOI Foundation

Norman Paskin

Kluwer Academic Publishers

Mike Casey

Library Binding Institute

J. Wesley Moyer

(continued)

NISO Voting Members *(continued)*

The Library Corporation

Mark P. Wilson
Nancy Capps (Alt)

Library of Congress

Winston Tabb
Sally H. McCallum (Alt)

Los Alamos National Laboratory

Richard E. Luce

Lucent Technologies

M.E. Brennan

Medical Library Association

Nadine P. Ellero
Carla J. Funk (Alt)

MINITEX

Cecelia Boone
William DeJohn (Alt)

Motion Picture Association of America

William M. Baker
Axel aus der Muhlen (Alt)

Music Library Association

Lenore Coral
Geraldine Ostrove (Alt)

National Agricultural Library

Pamela Q. J. Andre
Gary K. McCone (Alt)

**National Archives and Records
Administration**

Alan Calmes

**National Federation of Abstracting and
Information Services**

Marion Harrell

National Library of Medicine

Betsy L. Humphreys

OASIS

Laura Walker

OCLC, Inc.

Donald J. Muccino

Openly Informatics, Inc.

Albert Simmonds
Eric Hellman (Alt)

**Recording Industry Association of
America**

Linda R. Bocchi
Michael Williams (Alt)

Research Libraries Group, Inc.

Kathleen Bales
Wayne Davison (Alt)

RoweCom, Inc.

Robert Boissy
Marilyn Geller (Alt)

SilverPlatter Information, Inc.

Marwan Sabbouh

SIRS, Inc.

Leonardo Lazo
Harry Kaplanian (Alt)

Society for Technical Communication

Annette Reilly
Kevin Burns (Alt)

Society of American Archivists

Lisa Weber

Special Libraries Association

Marjorie Hlava

SUNY/OCLC Network

Mary-Alice Lynch
Jane Neale (Alt)

Triangle Research Libraries Network

Jordan M. Scepanski
Mona C. Coutts (Alt)

**U.S. Department of Commerce, National
Institute of Standards and Technology,
Office of Information Services**

**U.S. Department of Defense, Defense Technical
Information Center**

Gretchen A. Schlag
Gopalakrishnan Nair (Alt)

**U.S. National Commission on Libraries and
Information Science**

Denise Davis

VTLS, Inc.

Vinod Chachra

The H. W. Wilson Company

George I. Lewicky
Ann Case (Alt)

(continued)

At the time NISO released this Draft Standard for Trial Use, the following individuals served on its Board of Directors:

NISO Board of Directors

Donald J. Muccino, Chair
OCLC
Online Computer Library Center, Inc.

Beverly P. Lynch, Vice-Chair/Chair-elect
University of California - Los Angeles

Joel H Baron, Immediate Past Chair
The New England Journal of Medicine

Jan Peterson, Treasurer
Dawson Information Services Group

Patricia R. Harris, Executive Director
National Information Standards Organization

Pieter S. H. Bolman
Academic Press

Priscilla Caplan, SDC Chair
Florida Center for Library Automation

Carl Grant
Ex Libris (USA), Inc.

Brian Green
Book Industry Communication

Richard E. Luce
Los Alamos National Laboratory

Deanna B. Marcum
Council on Library and Information Resources

Norman Paskin
The International DOI Foundation

Steven Puglia
U.S. National Archives and Records Administration

Jordan M. Scepanski
Triangle Research Libraries Network

Albert Simmonds
Openly Informatics, Inc.

Standards Committee AP

The following individuals served on Standards Committee AP, the Book Item and Component Identifier Committee:

E Julia Blixrud, Committee Chair
Association of Research Libraries

D. Jeffrey Blumenthal
Follett Corporation

Tanny L. Franco
Defense Technical Information Center

Brian Green
Book Industry Communication

Ted Koppel
OCLC

Clifford Lynch
Coalition for Networked Information

Cliff Morgan
John Wiley and Sons

Mark H. Needleman
Data Research Associates, Inc.

Cecilia Preston
Preston & Lynch

Albert Simmonds
OCLC

Book Item and Component Identifier

1. INTRODUCTION

1.1 PURPOSE

This document defines the requirements for a variable length code that uniquely identifies each item or component (e.g., chapter, entry, illustration) within a book or publication to which an International Standard Book Number (ISBN) has been assigned. The acronym "BICI" stands for Book Item and Component Identifier, and is used in here to refer to the code itself.

1.2 SCOPE

The BICI is applicable to any book identified by an ISBN. Subject to the rules governing ISBN assignment, there is no restriction on the medium in which the publication may have appeared. (See Appendix A for more information about ISBN assignments.)

Within the set of books that meet these criteria, the BICI can be used to identify components such as:

- A part, a chapter or a section within a chapter, a volume within a set, or any other text component (e.g., an introduction, foreword, afterword, bibliography, index) that comprises an identifiable unit.
- An illustration, map, figure, table or other content that is a part of the publication.
- An entry or article that is separately enumerated in a publication such as a directory, encyclopedia, etc. that is not structured into chapters.

The BICI is intended to be created and used by those members of the community engaged in functions associated with the book, such as online retrieval, database linking, document delivery, rights management etc. Appendix E contains scenarios describing the use of the BICI in some of these functions.

1.3 PRINCIPLES AND GUIDELINES

1.3.1 Use of ISBN.

The book must have an ISBN before a BICI can be assigned to any of its components.

1.3.2 Self-derivability.

With access to a book or a citation and the necessary data elements, the assignment of a unique and unambiguous BICI to any component can be created by following a simple set of rules.

1.3.3 Use of all available data and redundancy.

The BICI should be constructed including all available data elements, even when some of them appear redundant, in order to ensure accurate identification.

1.3.4 Null data elements.

Certain data elements may be null. The code requires use of the ISBN, the delimiters for the *Component Segment*, and all elements in the *Control Segment*.

1.4 STRUCTURAL MODEL

The BICI is a combination of three segments, all of which are required. Interpretation is determined by the *Control Segment*. The three segments are:

Item Segment: the data elements used to identify an item. These data elements are:

- ISBN,
- Chronology, and
- Item enumeration.

Component Segment: the data elements needed to identify a component within an ISBN identified publication. These data elements are:

- Component enumeration,
- Title code and
- Location (i.e., pagination) or other numbering schema.

Control Segment: the data elements required to record the administrative elements that specify the validity, format, and version of the code representation.

- Code Structure Identifier (CSI),
- Component Type Identifier (CTI),
- Medium/Format Identifier (MFI),
- Standard Version Number (SVN) of BICI in use, and
- Check character.

Figure 1: Block diagram of the BICI structural model

Item Segment	Component Segment	Control Segment
ISBN(Chronology)Enumeration	(Enumeration; Title Code; Location or other numbering schema)	CSI:CTI:MFI;SVN-Check Character

2. REFERENCED STANDARDS

This standard is intended for use in conjunction with the following standards. When these standards are superseded by a revision approved by the American National Standards Institute, consult the revision.

ANSI/NISO/ISO 2108-1992, International Standard Book Numbering (ISBN).

ANSI X3.4-1986, American National Standard for Information Systems – Coded Character Sets – 7-Bit American National Standard Code for Information Interchange (7-bit ASCII).

3. DEFINITIONS

Book – Any publication to which an ISBN has been assigned in accordance with the ISBN rules.

Caption – The descriptive text for an illustration, photograph, diagram, table, etc. (see also Title). The caption follows any labels (e.g., “Figure 1 Three models for representation” – the caption begins with “Three”).

Character set – A collection of letters, digits, punctuation marks or other symbols. Such as ISO Latin-1.

Check Character – An alphanumeric character calculated from the BICI string, used to ensure consistency or validity of the BICI when it is recorded or transported across applications (see Section 6.3.5).

Chronology – Date of publication. May consist of just the year of publication, but may also indicate month, season, quarter etc., often used with regularly updated publications such as loose-leaf titles.

CIP – Cataloging in Publication, a program of the Library of Congress to provide descriptive bibliographic data for books prior to publication. The data is often found on either the title page verso or the last page of the book and available also through online databases.

Code Structure Identifier (CSI) – A code to distinguish the BICI type being constructed (see Section 6.3.1).

Component – Any identifiable portion of a book.

Component Segment – The group of data elements in the BICI string that identifies a specific portion of a book (see Section 6.2).

Component Type Identifier (CTI) – A code to distinguish a specific identifiable portion of a book (see Section 6.3.2).

Control Segment – The group of data elements in the BICI string that defines the structure and use of the code (see Section 6.3).

Book Item and Component Identifier

CSI – see Code Structure Identifier.

CTI – see Component Type Identifier.

Date – The year of publication, distribution, etc. of a book or the year (and month and day, if applicable) of publication of a component.

Enumeration – A scheme used by a publisher to identify a book or component e.g., volume designation in a multi-volume set, a chapter number, a part number.

International Standard Book Number (ISBN) – A number that uniquely identifies a book or publication. The ISBN appears in the BICI as a ten-character string without spaces or punctuation; the first nine characters are always numeric; the tenth character may be a numeric digit or the letter X, which represents the numeric value 10.

Item – An identifiable unit e.g., a book, a volume in a multi-volume set; etc. to which a BICI can be assigned.

Item Segment – The data elements in the BICI string that identify an item to which a BICI can be assigned. (See Section 6.1).

Hexadecimal (%) encoding – substituting for the given character its hexadecimal value preceded by percent (%) # becomes %23 for example.

Location – The scheme used to identify the placement of a component within a publication.

Medium/Format Identifier (MFI) – A two-letter alphabetic code used to identify the medium or format in which a publication is presented (see Section 6.3.3).

MFI – see Medium/Format Identifier.

Range – Beginning and ending location, e.g., page range.

Segment – A portion of the BICI code that is comprised of a set of related data elements.

Set – Two or more physical or logical volumes.

Title – The identifying name of a *text* component within a publication (e.g., chapter, section or encyclopedia entry). (See also Caption).

Title Code – A data element within the BICI component segment, derived from a title or caption, which serves as identification (see Section 6.2.2).

Version number – A number that identifies the version of the standard used to construct the BICI (see Section 6.3.4).

Volume – A single physical or logical unit of publication, a book to which an ISBN has been assigned.

4. CHARACTER SET

Transcribe alphabetic characters as they appear on the item. Encode alphabetic characters in upper case. Transliterate non-roman alphabetic characters to roman script, using the appropriate ANSI transliteration standards, or, if no ANSI standard is available, the ISO transliteration standards (See Appendix B).

Table 1 lists the *only* characters that are used in the construction of the BICI. Those characters in bold and italic require % encoding in order to be successfully parsed using HTTP, XML standards. (see Appendix F). Otherwise, these characters are a subset of the 7-bit ASCII character set and can be reliably used across various network and electronic mail systems. The individual data elements may impose additional constraints on allowable characters as illustrated in Table 2.

Table 1: Characters allowed in BICI ¹⁷

0-9 digits zero to nine	\$ dollar sign	+ plus sign
A-Z capital letters A to Z	= equals	? question mark
& ampersand	! exclamation mark	" quotation mark
' apostrophe	> greater-than sign (closing angle bracket)	; semicolon
* asterisk	- hyphen (minus)	/ forward slash
\ backslash	< less-than sign (opening angle bracket)	^ circumflex
} closing curly bracket	# number sign	' grave
) closing parenthesis	{ opening curly bracket	~ tilde
] closing square bracket	(Opening parenthesis	_ underscore
: colon	[opening square bracket	 vertical bar (fill)
, comma	% percent sign	
@ commercial at	. period	

Table 2: Legal character usage by data element

Data element	Length	Alpha	Numeric	Alpha- numeric	Other legal characters
Item segment					
ISBN	variable ¹⁸			√	X
Chronology	variable		√		
Item enumeration	variable			√	
Component segment					
Component enumeration	variable			√	
Title code	up to 6			√	
Location or Proprietary ID	variable			√	
Control segment					
Code structure identifier	variable		√		
Component type identifier	variable		√		
Medium/format identifier	variable	√			
BICI version number	1		√		
Check character	1			√	*

¹⁷ Characters in bold italic require hex encoding in URIs, see Appendix F.

¹⁸ The ISBN community is as of this writing considering structural changes and expansion of the code to 13 characters in congruence with Bookland EAN.

5. PUNCTUATION

Punctuation marks are used for structural identification and as demarcation symbols for distinguishing among data elements in the BICI.

Table 3: Use of Punctuation in BICI.

Symbol	Name	Purpose and example
()	Parentheses	In the Item Segment, to enclose the chronology. In the Component Segment to offset the segment.
:	Colon	In the Component Segment, between levels of structural hierarchy in the enumeration.
-	Hyphen	In the Component Segment, to link a first and last designator in a range (e.g. pages 123-125 or other ranges of location indicators). In the Control Segment, to separate the check character from the rest of the BICI string (e.g. 1.1.TX;1-X).
.	Period	In the Component Segment as used in the enumeration, location, range by the publisher (e.g., section 1.3). In the Control Segment, to separate each of the first three data elements (e.g. 1.1.TX;1-X).
+	Plus	In the Component Segment enumeration, to denote an insert (e.g. an illustration, a plate tipped in between sequentially numbered pages) that appears after the cited page number (e.g. 124+).
/	Forward slash	In the Chronology, as the connector that links the beginning and ending dates in a range (e.g., 1998/1999) In the Location, when specifying a range and the page is marked 3-1, (e.g. 3-1/3-15)
;	Semi-colon	In the Component Segment, between the Enumeration, Title Code, and the Location or other numbering scheme. In the Control Segment, to separate the medium/format identifier and the standard version number (e.g. 1.1.TX;1-X).
*	Asterisk	In the Control Segment, as a check character corresponding to the value 36 (see Appendix A).

6. SEGMENTS AND DATA ELEMENTS

The BICI code is made up of three segments: the *Item*, *Component* and *Control Segments*. This section defines the usage and format of the data elements within each segment.

The *Item Segment* is located at the beginning of the BICI code string. The ISBN is required. If needed to specify part of a set, chronology and/or enumeration of individual volumes may be provided. The chronology is enclosed in a set of parentheses followed by the enumeration data. Enumeration may be a volume number in a multi-volume set with one ISBN for the entire set. If the volume to be encoded within the multi-volume uses only words to distinguish the contents, than a BICI can not be derived for this work. For example a volume in an encyclopedia set with the enumeration of A – Axis, or a set of book in which each volume is on a specific subtopic where the enumeration is that topic (Ducks and Geese). (see Section 6.1)

The *Component Segment* is enclosed with parentheses (). It follows the *Item Segment*, and identifies a specific component of the book item by using a defined set of data elements and structural punctuation. These data elements include component enumeration, title data, location, range data or a publisher's own coding system. (see Section 6.2)

The *Control Segment* follows the *Component Segment*, and is administrative in function. By using the assigned codes, the *Control Segment* identifies which code structure is being used, the nature of the component that the BICI identifies, the medium or format of the item or component, the version of the standard used in creating the BICI, and a check character to verify the integrity of the BICI. (see Section 6.3)

Table 4: Relationships of Code Structure Identifiers (CSI) to other BICI segments

If Code Structure ID	Item Segment	Component Segment	Control Segment
1	ISBN(Chronology)Enumeration	()	1.CTI.MFI;SVN-Check Character
2	ISBN(Chronology)Enumeration	(Enumeration; Title Code; Location)	2.CTI.MFI;SVN-Check Character
3	ISBN(Chronology)Enumeration	Title Code; Other numbering schema)	3.CTI.MFI;SVN-Check Character

To illustrate: the BICI used to identify Chapter 10 "English as a world language" in *The English Language a Historical Introduction*, 1993 pp. 234-261 (ISBN 0-521-41620-5), would be **0521416205(1993)(10;EAAWL;234-261)2.2.TX;1-H**. Where the,

- *Item Segment* (see Section 6.1) consists of the string, **0521416205(1993)** which is the ISBN, the chronology 1993 in parentheses and no enumeration data,
- *Component Segment* (see Section 6.2) is the data contained in the second set of parentheses, **(10;EAAWL;234-261)**, and represents chapter 10 with a title code derived from the initial characters of the words in the title, which is on pages 234-261 and the
- *Control Segment* (see Section 6.3) follows the *Component Segment* and identifies a number of attributes of the BICI. In the example above, **2.2.TX;1-H** denotes the item being identified as a component (**2**), of the component type chapter (**2**), in print on paper (**TX**), using version (**1**) of the standard and a check character with the value (**H**).

6.1 ITEM SEGMENT

The *Item Segment* is comprised of the ISBN, chronology, and enumeration data elements. The ISBN is followed by the chronology that is enclosed in parentheses and followed by the item enumeration. Of the three data elements in the *Item Segment*, the ISBN is mandatory and the chronology and enumeration elements may be null.

Book Item and Component Identifier

To illustrate using the example in Section 6 above, if the only item level data available is the ISBN, then this segment would be encoded as **0521416205()**. Where the **0521416205** is the ISBN without hyphens and encoded as a simple 10 character string, and there is a null **()** chronology, and no enumeration. If all data elements were available, the item segment would be coded as **0262590107(1980)1**. **0262590107** is the ISBN, and the chronology is **(1980)** and the enumeration is **1** for volume 1 of the set.

6.1.1 ISBN

The ISBN for the book is recorded as a ten-character code without internal punctuation or spacing. Select the ISBN for the most specific level (e.g., the ISBN for an individual volume rather than for a set). In cases where there could be ambiguity in the selection of the ISBN, the following rules should apply:

Selecting an ISBN from a volume in hand:

(a) If there is only one ISBN on the title page or its verso.

Seuss-isms for Success: Insider Tips on Economic Health from the Good Doctor.
New York; Random House, 1999. From the title page ISBN: 0-679-89477-2.
BICI: **0679894772(1999)()1.0.TX;1-O**

McKeeman, W.M., J.J. Horning and D.B. Wortman. *A Compiler Generator.*
Englewood Cliffs, NJ; Prentice-Hall, 1970. From title page verso: 13-155077-2
Add the implied 0 and record.
BICI: **0131550772(1970)()1.0.TX;1-J**

(b) If there is more than one ISBN on the title page verso, and one ISBN is identified as relating to the format in hand (e.g., paperback, hardback), use it in preference to any other.

Bettig, Ronald *Copyrighting Culture: The Political Economy of Intellectual Property*
Boulder, CO; Westview Press, 1996. From CIP data: ISBN 0-8133-1385-6—
ISBN 0-8133-3304-0 (pbk)
BICI for the paperback: **0813333040(1996)()1.0.TX;1-P**
BICI for the hardback: **0813313856(1996)()1.0.TX;1-L**

(c) If there is more than one ISBN on the title page verso, use the one given that applies to the volume in hand. This may be indicated by the ISBN on the cover or jacket or other source.

Salton, Gerard and Hans-Jochen Schneider, ed. *Research and Development in Information Retrieval: Proceedings*, Berlin, May 18-20, 1982. New York; Springer-Verlag, 1983.

From title page verso: ISBN 3-540-11978-7 Springer-Verlag Berlin Heidelberg New York,
ISBN 0-387-11978-7 Springer-Verlag New York Heidelberg Berlin

From back cover: ISBN 3-540-11978-7, ISBN 0-387-11978-7

The ISBN used is from the Cataloging in Publication data often found on the verso of the title page for books distributed in the United States. The selection of ISBN in situations such as this will depend on local customs. This is an ANSI/NISO standard

and as such will use the local ISBN, as determined by cataloging rules.

BICI: **0387119787**(1983)01.0.TX;1-Y

- (d) If there is no other way of determining which of more than one ISBN is assigned to the book in hand, use the first ISBN on the title page verso or, failing that, on the cover or jacket.

MacWilliams, F.J. and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. New York; North-Holland, 1977.

From title page verso: ISBN 0 444 85009 0 and 0 444 85010 4

Appears on back cover: 0 444 85193 3

BICI: **0444850090**(1977)01.0.TX;1-2

- (e) When constructing a BICI without an item in hand, and if the ISBN is not known, choose the ISBN from an authoritative source such as a publisher or library catalog.

From a citation: Van Rijsbergen, C.J. *Information Retrieval*. Boston; Butterworths, 1979.

From Library of Congress bibliographic record web search: ISBN 0-408-70929-4

BICI: **0408709294**(1979)01.0.TX;1-4

Multi-volume publications:

- (a) If a multi-volume publication has a separate ISBN for each volume, use the volume level ISBN. In this case, the enumeration data element should be omitted since the volume ISBN identifies the item.

Williams, Martha E. ed., *Annual Review of Information Science and Technology*. Volume 33. Medford, NJ: Information Today, Inc., 1998.

From title page verso: 1-57387-065-X

BICI: **157387065X**(1998)01.0.TX;1-R

- (b) If a multi-volume publication does not have a separate ISBN for each volume, use the ISBN for the publication as a whole, and use any chronology and enumeration labels or data to identify the volume.

Iyanaga, Shōkichi and Yukiyoji Kawada. *Encyclopedic Dictionary of Mathematics*, Volume 1. Cambridge, MA; MIT Press, 1980.

From title page verso: 0-262-09016-3 (hard) 0-262-59010-7 (paper)

BICI: **0262590107**(1980)101.0.TX;1-Y (item in hand is the paperback)

6.1.2 CHRONOLOGY

The chronology is used both for the date of publication and for publications for which updates are issued from time to time, such as loose-leaf editions. Use chronology when available. Such usage is required for loose-leaf updates where the entire publication is usually identified by a single ISBN, yet the content of an individual leaf is specified by the date of the update.

Table 5: Chronology codes

Months		Seasons	Quarters
01 = January	07 = July	21 = Spring	31 = First quarter
02 = February	08 = August	22 = Summer	32 = Second quarter
03 = March	09 = September	23 = Fall	33 = Third quarter
04 = April	10 = October	24 = Winter	34 = Fourth quarter
05 = May	11 = November		
06 = June	12 = December		

The chronology field may take any of the forms YYYY (year), YYYYMM (year and month), YYYYMMDD (year, month and day). Combined chronology may be shown by using one of the forms YYYY/YYYY, YYYYMM/YYYYMM or YYYYMMDD/YYYYMMDD.

Darling, Pamela, with Duane E. Webster. *Preservation Planning Program: An Assisted Self-Study Manual for Libraries*. Washington, DC: Association of Research Libraries. Revised 1993 ed. ISBN 0-918006-69-4
 BICI: 0918006694(1993)01.0.TX;1-7

Mark, Zella. §5.131 Form: Affidavit to Verify Complaint in *California Paralegal's Guide*. 3rd ed. Volume 1. Los Angeles; Parker & Sons. 1989. (rev. 2-1989)
 ISBN: 0-911110-96-8.
 BICI: 0911110968(2-1989)1(5.131;FATVC;)2.2.TX;1-N

6.1.3 ITEM ENUMERATION

When an item is a volume of a multi-volume work and does not have its own unique ISBN, record the volume identifier as given on the piece. Alphabetic characters used in volume enumeration should be transcribed as they appear on the piece and encoded in upper case. Convert any word designations to their numerical equivalents (e.g., volume "Thirty-four" should be transcribed as "34"). When the volume designation is non-numeric, i.e. "Ducks and Geese" a BICI can not be assigned. If the volume is designated Vol. 4: Ducks and Geese the BICI can be constructed using the volume number "4".

Krause, Chester L. and Clifford Mishler. *Standard Catalog of World Coins*. Vol.2. Iola, WI: Krause Publications, 1991. 19th ed. 0-87341-151-X
 BICI: 087341151X(1991)201.0.TX;1-G

In a multi-volume set when individual volumes have not each been assigned an ISBN, use the ISBN for the set and the volume enumeration to create a BICI. Enumeration may be a volume number in a multi-volume set with one ISBN for the entire set.

CQ Almanac Volume XXXVII. Washington, DC; Congressional Quarterly, 1981.
 ISBN: 0-87187-231-5.
 BICI: 0871872315(1981)XXXVII01.0.TX;1-4

If the physical item carries two or more numbered volumes of the work in question, and this is the only enumeration, the first and last volume numbers, separated by a hyphen (-), should be recorded, e.g. 2-3.

6.2 COMPONENT SEGMENT

The *Component Segment* contains those data elements by which the component can be identified. This segment follows the *Item Segment* and is enclosed in parentheses. The data elements in this segment are the component enumeration, title code and location or proprietary identifiers. The component enumeration may consist of a chapter or section identifier; location data may include a starting page or page range; and private schemes such as a publisher's filename.

6.2.1 Enumeration

The data found in the component enumeration is specified in the Code Structure Identifier (see Section 6.3.1).

6.2.1.1 Entire volume. If the Code Structure Identifier (CSI) is 1, the BICI refers to the entire volume. When the BICI is used to identify a book item, the component segment is null, but must be present in the form of an empty pair of parentheses (). (see Section 6.1.3 for additional discussion of this component type.

Tanenbaum, Andrew S. *Computer Networks*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1981. ISBN 0-13-165183-8
BICI: 0131651838(1981)()1.0.TX;1-O

Iyanaga, Shōkichi and Yuklyosi Kawada. Volume 1 of *Encyclopedic Dictionary of Mathematics*. Cambridge, MA; MIT Press, 1980. ISBN: 0-262-59010-7
BICI: 0262590107(1980)1()1.0.TX;1-Y

6.2.1.2 If the CSI is 2, the BICI refers to an enumerated component part, that is, chapter, section, plate, etc., or a location. If the component is numbered (e.g. with a part, chapter, section number, or a figure or plate number), the number or numerals should be entered into the component enumeration field. Include any letters forming part of the numbering scheme (including Roman numerals) as they appear on the item and convert to uppercase.

A chapter:

Chiou-sen, Dora Chen. **Chapter 4**: Serials Vendors in *Serials management: a practical guide*. pp. **54-64**. Chicago; American Library Association, 1995. From the Cataloging in Publication Data ISBN 0-8389-0658-3
BICI: 0838906583(1995)(4;SV;54-64)2.2.TX;1-J

Multiple enumeration schemes are used in a single volume:

King, Stephen. **Book II**: On the border *The Stand*. Garden City, NY; Doubleday, 1978. pp. **[267]-622**. ISBN: 0-285-12168-7
BICI: 0285121687(1978)(II;OTB;267-622)2.2.TX;1-X

King, Stephen. **Chapter 34**. [untitled] In *The Stand*. Garden City, NY; Doubleday, 1978. pp. **269-285**. ISBN: 0-285-12168-7
BICI: 0285121687(1978)(34;;267-285)2.2.TX;1-V

Tables, graphs, charts, and other illustrative materials:

Pearson, Anna **Chapter 8**, Chart A in *The Complete Needlepoint Course*. Radnor, PA; Chilton, 1991. ISBN: 0-8019-8227-8
BICI: 0801982278(1991)(8;A;)2.3.TX;1-V

Book Item and Component Identifier

6.2.1.2.1 Hierarchy in Component Enumeration. The identification of specific components may be ambiguous. For example there may be a "Section 1" in each chapter of a text. Thus, the combination of Chapter and Section provides the specificity needed to construct the BICI code. The colon is used to separate levels in this hierarchy if not otherwise provided in the text.

Johnson, Norman L. and Samuel Kotz. **Chapter 4: Poisson Distribution Section 1:** Definition and Genesis in *Discrete distributions*. New York; John Wiley & Sons, 1969. ISBN: 0-471-44360-3
BICI: 0471443603(1969)(4:1;DAG;)2.2.TX;1-Y

In many technical publications the hierarchy is included in the numbering scheme, and as such is recorded as found.

Arnold, Ken and Bryan O'Sullivan, Robert W. Scheifler, Jim Waldo and Ann Wollrath. **AR.1.1:** Goals of the System. *The Jini™ Specification*. Reading, MA; Addison-Wesley, 1999. ISBN: 0-201-61634-3
BICI: 0201616343(1999)(AR.1.1;GOTS;)2.2.TX;1-B

6.2.1.2.2 Location information is used in addition to the enumeration data to identify the component when available; for example, by a page number (start of component) or page range. This is often the only form of identification for the start of a work that is published in a proceedings for example.

Bauer, Charles, [Chapter] 2 The flux of language. In *The English Language a Historical Introduction*, Cambridge; Cambridge University Press, 1993. pp. **32 – 57** ISBN: 0-521-42622-7
BICI: 0521426227(1993)(2;TFOL;**32-57**)2.2.TX;1-8

Cooper, William S., "Getting Beyond Boole" In Readings in *Information Retrieval* edited by Karen Spark Jones and Peter Willett. **p. 265** San Francisco; Morgan Kaufmann, 1997. ISBN from Cataloging in Publication Data: 1-55860-454-5.
BICI: 1558604545(1997)(;GGB;**265**)2.6.TX;1-H

When location is not printed on the page, it may be determined from the adjacent pages. Often illustrations or plates are inserted (tipped in) between two sequentially numbered pages. To identify this material, use the preceding page number followed by a plus sign (e.g., 174+).

6.2.1.3 A CSI of 3 indicates that a proprietary identification system is in use.

Alternative component identification schemes used primarily by publishers in editorial and electronic publishing processes may be incorporated in the BICI. When these schemes are present, the Code Structure Identifier equals 3.

"ABO Blood Group System" from *Encyclopedia of Immunology* 2nd ed. online pub. by Academic Press. DOI 10.1006/rwei.1999.0001 ISBN: 0122267656 (access to the online is included with the purchase of the print)
BICI: 0122267656(1999)(;ABGS;10.1006/rwei.1999.0001)**3**.2.CO;1-T

NOTE: in the CSI-3 the code carried in the *Component Segment* uses the syntax of the alternative identification scheme. In the example above the DOI is case sensitive and is recorded according to the DOI rules.

The BICI can also be carried in other identifier syntaxes; for example, the BICI above can be embedded in an URN as follows:

URN:BICI:0122267656(1999)(; ABGS;10.1006/rwei.1999.0001)2.2.CO;1-T

6.2.2 Title Code

A title code should be derived from the title of a component (such as the name of a chapter, the caption to an illustration or table). Encoded in uppercase in accordance with the following rules:

(a) Subject to the rules given in (b) through (g) below, use any and all title words without attempting to distinguish titles from subtitles.

(b) A word in a title or caption is defined as any character string delimited by spaces.

A Note on the Methodology of Information UK 2000 there are nine words in this title.

(c) Ignore accents, diacritics and punctuation marks. A punctuation mark that is not bound to any other characters – e.g. a dash as in this sentence – is to be ignored, however, and does not constitute a ‘word’.

File Design for On-Line Systems there are five words in this title. “On-line” is one word as the – is ignored.

It Takes “Twos” to Tangle contains five titles words. The quotation marks are ignored.

(d) The title begins after any labels: Chapter, Section, Figure, Table, Appendix, etc.

Appendix A: Critical Dates in History the title code is constructed from *Critical Dates in History*. If there is no descriptive title only a label as in *Appendix A* use *A* in the title code.

(e) Select the first six words as defined (or fewer if the title or caption has less than six words).

A Note on the Methodology of Information UK 2000 only *A Note on the Methodology of* would be used in the construction of the title code.

File Design for On-Line Systems is only five words long (see (c) above) and the words used to construct the title code are *File Design for OnLine Systems*

(f) Do not use authorship or other related information as part of the title, except when an author’s name is given within the title.

(g) Transliterate any word that is in a non-roman alphabet.¹⁹ Convert any mathematical, scientific or other symbol that is not part of the set listed in Table 1, to the English word or words describing the symbol, without an initial article, and without spaces. This process of transliterating any non-alphanumeric symbols will in no circumstances have the effect of changing the word boundaries identified in step (b). Treat roman numerals as “words”.

Some additional examples of title codes.

Section title examples:

I. **Research at Short and Ultrashort Wavelengths**, in Chapter 5. In: *AT&T Bell Laboratories. A history of engineering and science in the Bell System: Communication sciences (1955-1980)*. S. Millman ed. [Short Hills, NJ ?]; AT&T Bell Laboratories, 1984

BICI: 0932764061(1984)(5:I;RASAUW;)2.2.TX;1-C

¹⁹ See Appendix B for relevant standards.

Book Item and Component Identifier

1.1 **Picture Scanning** in Chapter 3, Picture Communication Research. In: *AT&T Bell Laboratories. A history of engineering and science in the Bell System: Communication sciences (1955-1980)*. S. Millman ed. [Short Hills, NJ ?]; AT&T Bell Laboratories, 1984
BICI: 0932764061(1984)(3:1.1;**PS**;)2.2.TX;1-F

Caption title examples:

Figure 3-11. **The first solid-state color television camera** (page 159 of chapter 3, Picture Communication Research) In: *AT&T Bell Laboratories. A history of engineering and science in the Bell System: Communication sciences (1955-1980)*. S. Millman ed. [Short Hills, NJ ?]; AT&T Bell Laboratories, 1984
BICI: 0932764061(1984)(3-11;**TFSCCTC**;159)3.3.TX;1-4

Graph 2: **TULIP Logfile Analysis: Georgia Tech: Usage per usertype per month**. In: *The TULIP Report*. New York; Elsevier Science, 1996. ISBN: 0-444-82540-1.
BICI: 0444825401(1996)(2;**TLAGTU**;)2.3.TX;1-L

Unnamed table on page 197. *The TULIP Report*. New York; Elsevier Science, 1996.
ISBN: 0-444-82540-1.
BICI: 0444825401(1996)(;;**197**)2.3.TX;1-*

6.3 CONTROL SEGMENT

The Control Segment specifies the type of data and level of identification for the BICI and provides interpretation of the BICI code. The control segment has five data elements, all of which must be present, with separators as defined below:

Code Structure Identifier(CSI)	Followed by period (.)
Component Type Identifier (CTI)	Followed by period (.)
Medium/Format Identifier (MFI)	Followed by semi-colon(;))
BICI Standard Version Number	Followed by hyphen (-)
Check Character	End of string no marks

Bauer, Charles, [Chapter] 2 The flux of language. In *The English Language a Historical Introduction*, Cambridge; Cambridge University Press, 1993. pp. 32 – 57 ISBN: 0-521-42622-7
BICI: 0521426227(1993)(2;TFOL;32-57)**2.2.TX;1-8**

6.3.1 Code Structure Identifier (CSI)

The Code Structure Identifier (CSI) is used to specify the type of data found in the Contribution Identifier. This standard recognizes four structurally different ways in which a component segment can be identified. These differences are specified in the Code Structure Identifier and are diagramed here to frame the discussion of the individual segments and the data elements contained in each. (Future modifications may bear other numbers).

Table 6: CSI types

CSI	Item Segment	Component Segment	Control Segment
1	ISBN(Chronology)Enumeration	()	CSI.CTI.MFI;SVN- Check Character
2	ISBN(Chronology)Enumeration	(Enumeration; TitleCode;Location)	CSI.CTI.MFI;SVN- Check Character
3	ISBN(Chronology)Enumeration	(;TitleCode; Proprietary Code)	CSI.CTI.MFI;SVN- Check Character

6.3.1.1 CSI = 1: Identifies an item or entire work.

Tanenbaum, Andrew S. *Computer Networks*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1981. ISBN 0-13-165183-8
BICI: 0131651838(1981)()1.0.TX;1-O

6.3.1.2 CSI = 2: Identifies a component of a book, e.g., book chapter or section, a graph or table, the preface, glossary or location.

King, Stephen. **Chapter 34** [untitled]. In *The Stand*. Garden City, NY; Doubleday; 1978. ISBN: 0-285-12168-7
BICI: 0285121687(1978)(34;;)2.2.TX;1-G

Cooper, William S., "Getting Beyond Boole" In Readings in *Information Retrieval* edited by Karen Spark Jones and Peter Willett. p. 265 San Francisco; Morgan Kaufmann, 1997. ISBN from Cataloging in Publication Data: 1-55860-454-5.
BICI: 1558604545(1997)(;GBB;265)2.6.TX;1-H

6.3.1.3 CSI = 3: Provides an encoding method for proprietary systems, such as the file names used in identifying a dictionary entry on CD-ROM. By identifying a BICI as Type 3, systems are alerted to the possibility that the data contained in the Component Segment enumeration data element may not be usable without prior knowledge of the encoding system.

6.3.2 Component type identifier (CTI)

Assign these values to the logical structured divisions of a book, as follows:

Table 7: CTI types

CTI	Description
0	Whole item
1	Front matter
2	Main text subdivision: Parts, Chapters, Sections, Acts etc.
3	Discrete object with in a book: Abstracts, Tables, Figures, etc.
4	Back matter: Index, Appendices, References, Bibliographies, etc.
5	Supplementary materials: Maps, Disks, Erratum, etc.

Book Item and Component Identifier

6.3.2.1 CTI = 0 The whole item. For example, a single volume in multi-volume set where the set has a single ISBN.

Iyanaga, Shōkichi and Yukkyōsi Kawada. *Encyclopedia Dictionary of Mathematics*. Volume 1 Cambridge, MA; MIT Press, 1980. ISBN: 0-262-59010-7
BICI: 0262590107(1980)1(0.TX;1-Y

6.3.2.1 CTI = 1 Front matter. That is, pages of a book that precedes the major subdivisions (such as chapters). Conventionally paged with small roman numerals, but not necessarily so. There are probably literally hundreds of things that could go into preliminaries such as: frontispiece, half-title page, series page, title page, title verso, Table of Contents, dedication, acknowledgments, preface to series, preface to first edition, preface to subsequent editions, foreword, Editorial Board, as well as, lists of contributors, figures, tables, symbols, abbreviations.

Bolter, Jay David. **Contents**. In *Writing space: The computer, hypertext, and the history of writing*. p. v. Hillsdale, NJ; Lawrence Erlbaum, 1991. ISBN: 0-8058-0428-5
BICI: 0805804285(1991)(;C;V)2.1.TX;1-S

Bolter, Jay David. **Preface**. In *Writing space: The computer, hypertext, and the history of writing*. p. ix-xi Hillsdale, NJ; Lawrence Erlbaum, 1991. ISBN: 0-8058-0428-5
BICI: 0805804285(1991)(;P;IX-XI)2.1.TX;1-*

Devlin, Keith. [**Title page**] *InfoSense: Turning information into knowledge*. p. [iii]. New York; W.H. Freeman and Company, 1999. ISBN: 0-7167-3484-2
BICI: 0716734842(1999)(;III)2.1.TX;1-A

6.3.2.2 CTI = 2 Main text subdivision(s) Used for Parts, Chapters, Sections, Acts (of plays), etc. - whatever the book uses to divide into discrete units. Again, there is no standard terminology, so one book may be divided up into Sections that include Chapters, while another could be divided up into Chapters that included Sections.

Berinstein, Paula. **Chapter 5** "How to Search for Images: Selecting a Place to Search" In: *Finding images online*. Wilton, CT; Pemberton Press; 1996. ISBN: 0910965218
BICI: 0910965218(1996)(5;HTSFIS;)2.2.TX;1-H

Part II: Heading, Uniform Titles, and References. In *Anglo-American cataloging rules*. Michael Gorman and Paul W. Winkler eds. 2nd ed. 1988. Chicago; American Library Association, 1988. ISBN: 0-8289-3360-2.
BICI: 0828933602(1988)(II;HUTAR;)2.2.TX;1-X

Chapter 24: Headings for Corporate Bodies. In *Anglo-American cataloging rules*. Michael Gorman and Paul W. Winkler eds. 2nd ed. 1988. Chicago; American Library Association, 1988. ISBN: 0-8289-3360-2.
BICI: 0828933602(1988)(24;HFCB;)2.2.TX;1-E

Chapter 24: Headings for Corporate Bodies, **Section 7B4**: Location In *Anglo-American cataloging rules*. Michael Gorman and Paul W. Winkler eds. 2nd ed. 1988. Chicago; American Library Association, 1988. ISBN: 0-8289-3360-2.
BICI: 08289336021988(1988)(24;7B4;L)2.2.TX;1-E

6.3.2.3 CTI = 3 Used for discrete objects with in a book. In print works: abstracts, tables, spreadsheet, worksheet, sidebars, mathematical formula, figures, diagrams, plates, three-dimensional pop-ups. In electronic works, a sound clip for example.

Edley, Joe and John D. Williams, Jr. **Diagram 14-4**. In *Everything SCRABBLE®*. New York; Pocket Books, 1994. pp. 137 ISBN: 0-671-86686-9
BICI: 0671866869(1994)(14-4;;137)2.3.TX;1-H

Matthews, Joseph L. and Dorothy Matthews Berman. Sidebar entitled Sign Up Three Months Before Your Birthday. In *Social security, Medicare, and pensions. Chapter 7 Section 3*. Berkeley, CA; Nolo Press, 1996. ISBN: 0-87337-289-1
 BICI: 0873372891(1996)(7:3;SUTMBY;)2.3.TX;1-T

Johnson, Norman L. and Samuel Kotz. **Equation (34) in Chapter 6, Section 8.6** In *Discrete distributions*. New York; John Wiley & Sons, 1969. ISBN: 0-471-44360-3
 BICI: 0471443603(1969)(6:8.6;;34)2.3.TX;1-2

Lacambre, Geneviève. **Figure 2**. Cham, "The Sphinx of M. Gustave Moreau preventing M. Courbet from sleeping," *Le Charivari* (1864). Paris, Musée Gustave Moreau, Inv. 14581. p. 82. In *Gustave Moreau: Between Epic and Dream*. Princeton, NJ; Princeton University Press, 1999. ISBN: 0-691-00734-9
 BICI: 0691007349(1999)(2;CTSOMG;82)2.3.TX;1-V

6.3.2.4 CTI = 4 Back Matter. Material that follows the main text. In printed books, often the index, appendices, references.

Appendix C: Sample Tables for Current Library Services in *Planning Process for Public Libraries*, Chicago, IL: American Library Association, 1980. Pp. 301-304. ISBN 0-8389-3246-0

BICI: 0838932460(1980)(C;STFCLS;)2.4.TX;1-4

"**Selected Bibliography**" in *A Planning Process for Public Libraries*, Chicago, IL: American Library Association, 1980. Pp. **301-304**. ISBN 0-8389-3246-0

BICI: 0838932460(1980)(;SB; 301-304)2.4.TX;1-V

6.3.2.5 CTI = 5 Supplementary material and packaging materials. Material that is supplied physically separate from the printed book. Such as, teacher guides, workbooks, maps, disks, CD-ROMs, erratum, corrigendum, jacket, and other packaging (e.g. a box)

6.3.3 Medium/Format Identifier (MFI)

The medium/format identifier allows distinctions to be made among a variety of presentation formats in which the components may be made available. Table 7 shows the codes to be used. These codes are a subset of the codes included in ANSI/NISO Z39.56, 1996.

Table 8: MFI codes

TX	Printed text	SC	Sound recording
CD	Computer-readable optical media	TB	Braille
CO	Online	VX	Video recording
CP	Computer-readable magnetic disk media	ZN	Multiple physical forms
CT	Computer-readable magnetic tape media	ZU	Physical form unspecified
HE	Microfiche	ZZ	Other physical form
HD	Microfilm		

6.3.4 Standard Version Number

The Standard Version Number identifies the version of the standard used to create the BICI. It is preceded by a semicolon (;) and followed by a hyphen (-). The BICI standard carries the version number in a prominent position on the title page. For a BICI code constructed in accordance with the present specification, the version number is 1.

6.3.5 Check Character

The check character is calculated by applying a Modulus 37 algorithm to the characters in the BICI string. The procedure for doing so is given in Appendix C.

Appendix A

ISBN Information

ISBN Numbers

According to the article "How to Obtain an ISBN," written by Emery Koltary, Director of the United States ISBN Agency, and published in the 1998 Bowker Annual: Library and Book Trade Almanac (pages 592-595), the International Standard Book Numbering (ISBN) system began in the United Kingdom in 1967 and was introduced into the United States by the R.R. Bowker Company in 1968. Its purpose is to coordinate and standardize the use of identifying numbers so that each ISBN is unique to a title, edition of a book, or monographic publication published or produced by a specific publisher or producer.

Each ISBN number consists of ten digits separated into the following parts:

- Group identifier (national, geographic, language, or other convenient group)
- Publisher or producer identifier
- Title identifier
- Check digit

Administration of the ISBN system is carried out through the International ISBN Agency in Berlin, Germany; the national agencies; and the publishing houses themselves. In the U.S., assignment of ISBNs is done by the U.S. ISBN Agency and the publisher. There are standards governing printing of ISBN numbers of materials and the printing of the ISBN in Machine-Readable Coding.

For an ISBN application form and additional information, contact the United States ISBN Agency, R.R. Bowker Company, 121 Chanlon Road, New Providence, NJ 07974 (telephone: 908-665-6770). The e-mail address is isbn-san@bowker.com. The ISBN web site is at <http://www.bowker.com/standards/>.

Because of the changes within the ISBN community watch the web site.

It should also be noted here the ISBN has been assigned to non-book items, for example, greeting cards, bookmarkers and other items which are not what one would consider a book or its equivalent in another format such as, microprint or online. The BICI is not intended to accommodate these materials.

Appendix B

Related Standards and References

ANSI/NISO Z39.47-1993, *Extended Latin Alphabet Coded Character Set for Bibliographic Use*.

ANSI/NISO Z39.56-1996 *Serial Item and Contribution Identifier*.

ANSI/NISO Z39.64-1989 (R1995), *East Asian Character Code for Bibliographic Use*.

ISO 9:1995 (E), *Documentation — Transliteration of Arabic characters into Latin characters*.

ISO 233-2:1993(E), *Documentation — Transliteration of Arabic characters into Latin characters. — Part 2: Simplified transliteration*.

ISO 259:1984(E), *Documentation — Transliteration of Hebrew characters into Latin characters*.

ISO 259-2:1994(E), *Information and documentation — Transliteration of Hebrew characters into Latin characters — Part 2: Simplified transliteration*.

ISO/R 843:1968(E), *International systems for the transliteration of Greek characters into Latin characters*.

ISO 3602:1989(E), *Documentation — Romanization of Japanese (kana script)*.

ISO 7098:1991(E), *Information and documentation — Romanization of Chinese*.

ISO 8859, *Information Processing – 8-bit single-byte coded graphic character sets*.
Consists of ten parts. Each part specifies a set of up to 191 graphic characters and the coded representation of each of these characters by means of a single 8-bit byte. The use of control functions for the coded representation of composite characters is prohibited by ISO 8859. Each set is intended for use for a group of languages. This set of graphic characters is suitable for use in a version of an 8-bit code according to ISO 2022 or ISO 4873.

ISO 8859-1:1987(E), *Latin alphabet no. 1*. For use with at least: Danish, Dutch, English, Faeroes, Finnish, French, German, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish, and Swedish.

ISO 8859-2:1987(E), *Latin alphabet no. 2*. For use with at least: Albanian, Croon, Czech, English, German, Hungarian, Polish, Romanian, Serbian, Slovak, and Slovene.

ISO 8859-3:1988(E), *Latin alphabet no. 3*. For use with at least: Afrikaans, Catalan, Dutch, English, Esperanto, German, Italian, Maltese, Spanish, and Turkish.

ISO 8859-4:1988(E), *Latin alphabet no. 4*. For use with at least Danish, English, Estonian, Finnish, German, Greenlandic Lappish, Latvian, Lithuanian, Swedish, and Norewegian.

ISO 8859-5:1988(E), *Latin/Cyrillic*. For use with at least Bulgarian, Belorussian, Croatian, English, Macedonian, Russian, Serbian, and Ukrainian.

ISO 8859-6:1987(E), *Latin/Arabic alphabet*.

ISO 8859-7:1987(E), *Latin/Greek alphabet*. Is suitable for multiple language applications involving Latin and Greek scripts.

ISO 8859-8:1988(E), *Latin. Hebrew alphabet*.

ISO 8859-9:1989(E), *Latin alphabet no. 5*. For use with at least: Danish, Dutch, English, Finnish, French, German, Irish, Italian, Norwegian, Portuguese, Spanish, and Turkish.

ISO 8859-10:1992(E), *Latin alphabet no.6*. For use with at least: Danish, English, Estonian, Finnish, German, Greenlandic, Icelandic, Sami (Lappish), Latvian, Lithuanian, Norwegian, Faroese, and Swedish.

Appendix C

Calculation of Modulus 37 Check Character

The use of a check character helps to guard against errors resulting from improper data transcription. The check digit used in the Book Item and Component Identifier is calculated on a Modulus 37 basis as indicated in Table C-1.

Table C-1: Check Character Values

Char/Value	Char/Value	Char/Value	Char/Value	Char/Value	Char/Value	Char/Value
0=0	5=5	A=10	F=15	K=20	P=25	V=31
1=1	6=6	B=11	G=16	L=21	Q=26	W=32
2=2	7=7	C=12	H=17	M=22	R=27	X=33
3=3	8=8	D=13	I=18	N=23	S=28	Y=34
4=4	9=9	E=14	J=19	O=24	T=29	Z=35
					U=30	*=36

The modulus value obtained by the check character calculation will be substituted for the corresponding number or capital letter. The value "36" will be represented by the symbol "*" (asterisk). All punctuation characters are included in the check character calculation (Table C-2).

Table C-2: Example of Check Character Calculation

Step	Example: 046509175X() (9;WOGI;)1.0.TX;1-
1. Use Table A-1, Check Character Values, to assign numeric values to each alpha character in the human-readable string.	[0][4][6][5][0][9][1][7][5][33][00][01][9][:][32][24][16][18][:][:][1][.]][0][.]][29][33][:][1][:]
2. Replace all remaining, non-alphabetic characters (e.g. punctuation marks) in the human-readable string with the numeric value of 36.	[0][4][6][5][0][9][1][7][5][33][36][36][36][9][36][32][24][16][18][36][36][1][36][0][36][29][33][36][1][36]
3. Starting from the right most position of the new all-numeric string, sum all of the odd position values. The right-most position (i.e. position 1) will always be the hyphen (value 36) that directly precedes the check character.	[0][4][6][5][0][9][1][7][5][33][36][36][36][9][36][32][24][16][18][36][36][1][36][0][36][29][33][36][1][36] 36+36+29+0+1+36+16+32+9+36+33+7+9+5+4=289
4. Multiply the sum obtained in step 3, by three.	289x 3=867
5. Sum the remaining values in the original string (the even position numbers) with the result of step 4.	0+6+0+1+5+36+36+36+24+18+36+36+36+33+1=304 304+867=1171
6. Integer divide the result of step 5 by 37.	1171/37=31 r30
7. Subtract the integer remainder from 37. The result is the modulus 37 check character value. If the division has a remainder of zero, the zero(0) is the check character.	37-30=7
8. Use Table A-1, Check Character Values to match the remainder value to the appropriate check character: Numeric, Alpha (uppercase) or asterisk (*).	7=7
9. Final BICI string	046509175X() (9;WOGI;)1.0.TX;1-7

Appendix D

Usage Scenarios

Abstracting and Indexing Services

A BICI code can be used by A&I services as data (a) stored directly in the database records provided by A&I database suppliers or (b) constructed dynamically based on data stored in database records loaded from the original data providers. This BICI code, either displayed to the end user or hidden behind the scenes, could be used by the A&I service when a user requests a copy of the associated document. The BICI could be used either by the A&I service itself to supply a document, or it could be passed along to a third party supplier if the A&I service is not able to supply the requested document.

Document Delivery Services

A BICI code can be one of the identifying elements sent to a Document Delivery service provider to identify the document desired. This transmission could be automatic by another online system in which the user had identified the document, or it could be done manually by the end user.

Coursepack Production

BICI codes can be used to identify the multiple items that are incorporated into student coursepacks.

Library Reserve Rooms

In academic settings when sets of selected readings are placed in separate areas or delivered throughout separate systems, the BICI can be used as a data collection element. It could also be used as the link between items and retrieval systems.

Rights Management Agencies

BICI codes could be used to track requests for permission to use items or contributions and as reporting elements between rights holders and rights management agencies.

DOI System

A BICI code can be the suffix element in the DOI syntax.

Appendix F

% (Hex) Encoding

Encoding Requirements

BICIs may be placed within URLs and transmitted via HTTP to be resolved and special encoding requirements are required. The syntax for Uniform Resource Identifiers (URIs) is much more restrictive than the syntax for the BICI. A URI can be a Uniform Resource Locator (URL) or a Uniform Resource Name (URN). There will be instances when certain characters contained in BICIs will need to be encoded for transmission within a URL.

Encoding Characters

Since URL syntax conventions are more restrictive than BICI conventions, hexadecimal (%) encoding must be used for characters in a BICI that are not allowed, or have other meanings, in URLs. Hex encoding, or escaping characters, consists of substituting for the given character its hexadecimal value preceded by percent. Thus, # becomes %23 and BICI [need example] is encoded as [need example]. The browser does not now encounter the bare #, which it would normally treat as the end of the URL and the start of a fragment.

[or example for *]

NOTE: The BICI itself does not change with encoding, merely its representation in a URL. A BICI that has been encoded is decoded before being transmitted. Generally, decoding is handled by proxy servers. BICI creators and maintainers need to keep the following rules in mind: There are few character restrictions for BICI strings per se. When BICIs are embedded in URLs, they must follow the URL syntax conventions, but the same BICIs need not follow those conventions in other contexts, e.g. inventory databases, bibliographic catalogs. Always present BICIs in URLs in their encoded form to end-users.

The percent character (%) must always be hex encoded (%25) in any web form or URL. Other excluded characters must always be hex encoded in URLs.

Mandatory/Recommended Encoding

Table 1 shows a list of characters that are not allowed within the URL syntax: these native characters must therefore be replaced by hex-encoding. Table 2 shows additional characters where it is recommended that characters be replaced by hex-encoding. The distinction between Tables 1 and 2 is between practical experience with current web browsers and the more formal specification of URL syntax. The web is still a relatively new environment and both the formal specifications and web browser implementations are something of a moving target. It can say with certainty that the characters in Table 1 must be hex encoded. It cannot say with certainty that the characters in Table 2 must or must not be hex encoded, only that the formal specifications recommend it but no recent set of browsers require it. The tables are formatted to show the given character,

the encoding required by the BICI for use within a URL and finally what is stored. The | symbol is used to show a series of acceptable alternatives, e.g., the fact that # can be entered into a web form as either # or %23 is shown as # | %23. Note that all the given characters must be hex encoded within URLs but that only %, because of its use as marking the start of a hex encoding, must be hex encoded.

Table F-1 – Mandatory Encoding

Character	URL
#	%23
%	%25
SP	%20
"	%22

Table F-2 – Recommended Encoding

Character	URL
<	%60
>	%62
{	%123
}	%125
^	%94
[%94
]	%93
'	%96
	%124
\	%92