

Danijela Getliher, Jasenka Zajec

National and University Library in Zagreb

IDENTIFIERS FOR DIGITAL HERITAGE

Abstract: The increasing number of digitised publications published by national libraries and other heritage institutions stresses the importance of identification and moves the development of traditional identification systems to the next level. An overview of identifiers used for digitised heritage is given. The development, usage and benefits of the use of identification systems like DOI (Digital Object Identifier), ISBN (International Standard Book Number), ISSN (International Standard Serial Number), ISTC (International Standard Text Code), NBN (National Bibliography Number), SICI (Serial Item and Contribution Identifier), URN (Uniform Resource Number) are presented in more detail. Possible levels of identification i.e. different levels of granularity that can be used in identification of digitised publications are presented.

Keywords: digital heritage, digitised publications, identifiers, DOI, ISBN, ISSN, ISTC, NBN, SICI, URN.

1. Introduction

In an effort to open up and increase access to their collections, heritage institutions have been increasingly digitising their materials (books, newspapers, journals, maps, printed music and so on) and making them available on the Web for browsing and downloading by researchers and other users. Digitisation and all subsequent activities, up to the actual publishing of the material on the Web, influenced in a lot of ways the traditional workflows in libraries and other heritage institutions. New rules and standards had to be observed in abiding by the basic principles: protection of rare and valuable originals and production of important and enduring resources that are easily and infallibly identifiable and accompanied by high quality metadata.

The need to identify different content and objects is becoming more and more important, especially since digitisation produces the constantly growing number of digital objects that can be slightly different, similar or completely identical. The use of persistent identifiers, many of which have been well established in the library world for decades, facilitates organisation of digitised objects and their reliable retrieval. But digitisation has also expanded the scope of the already existing identifiers.

In this paper we shall try to address the basic issues of identification of digitised objects, give an overview of identifiers for digitised heritage and present the development, usage and benefits of their use. The most often used identifiers will be presented in more detail: DOI (Digital Object Identifier), ISBN (International Standard Book Number), ISSN (International Standard Serial Number), ISTC (International Standard Text Code), NBN (National Bibliography Number), SICI (Serial Item and Contribution Identifier) and URN (Uniform Resource Number).

2. What is an identifier?

An identifier is a name that serves to identify either a unique object or a unique class of objects, where the "object" or class may be an idea, physical or virtual object or substance. Identifier may be a word, number, letter, symbol, or any combination of these [10]. Identifiers must be linked to metadata

describing the identified digitised object to enable precise identification, access, transfer of metadata, distinguishing between different objects and reliable retrieval, but also to facilitate the transmission of data, citations, and links to other systems, increase availability and awareness of an object's existence. Identifiers bring stability and standardisation in digitisation processes and thereby can be used to reduce duplication.

Given their importance, there are a great number of different identifiers and new ones are being developed all the time so the following list is by no means exhaustive:

- DOI (Digital Object Identifier)
- ISAN (International Standard Audio-visual Number)
- ISBN (International Standard Book Number)
- ISCI (International Standard Collection Identifier)
- ISIL (International Standard Identifier for Libraries and Related Organizations)
- ISMN (International Standard Music Number)
- ISNI (International Standard Name Identifier)
- ISRC (International Standard Recording Code)
- ISSN (International Standard Serial Number)
- ISTC (International Standard Text Code)
- ISWC (International Standard Musical Work Code)
- NBN (National Bibliography Number)
- SICI (International Standard Collection Identifier)
- URN (Uniform Resource Name)

The identifiers are unique, i.e. they refer only to the object to which they were assigned, they are persistent in the sense that they refer to the object permanently. Identifiers can be dumb i.e. they have no meaning (like for instance ISSN that is made of eight digits) or intelligent (like ISBN that includes information about identified object: country, publisher, publication). Some identifiers were developed for printed resources but have been adapted to electronic and digital resources (ISBN and ISSN for instance) and some were developed as the result of intensively growing number of digital resources (like URN). However, they are not competitors but compatible, and some identifiers are interoperable like for instance embedding of ISSN or ISBN in URN [8] or in DOI. Actually, the first attempt of URN-ISSN interoperability was done in 2001 [9].

Identification is possible at different levels of granularity and identifiers can be assigned to different resources (e.g. books, articles, journals). Identifiers are assigned by different actors in the digitisation chain (by authorised agencies or by users themselves - like authors, publishers, libraries, and so on).

The assignment and use of the majority of identifiers is based on international standards and many of them especially the traditional ones like ISBN, ISSN are linked to ISO (International Standardisation Organisation) and those that are actionable on the Internet, like URN to the IETF (Internet Engineering Task Force). One of the differences is that the Internet standards are changed more easily, while ISO procedures are slightly more formal and complex.

3. Identifiers for Digitised Resources

The majority of identifiers that have been used in bibliographic communities for decades are well adapted to digitised resources. The following identifiers that are used for digitised materials will be presented in more detail: DOI (Digital Object Identifier), ISBN (International Standard Book Number), ISSN (International Standard Serial Number), ISTC (International Standard Text Code),

NBN (National Bibliography Number), SICI (Serial Item and Contribution Identifier), URN (Uniform Resource Number). Some of their basic characteristics are listed in Table 1.

Table 1: Identifiers and their characteristics

Identifier	Scope	Established	Published standard	Maintenance web address	U*	P* ¹	R* ²	M* ³	Structure/Syntax, example
DOI (Digital Object Identifier)	objects of any material form	1997	ISO 26324:2012	International DOI Foundation http://www.doi.org	+	+	+	+	prefix (directory, registrant) and suffix element, no limit on length doi:10.1006/jmbi.1998.2354
ISBN (International Standard Book Number)	monographic publications	1967	ISO 2108:2005	International ISBN Agency http://isbn-international.org/	+	+	+	+	prefix, registration group, registrant, publication, check digit 13 digits ISBN 978-953-304-412-5
ISSN (International Standard Serial Number)	serials and other continuing resources	1970	ISO 3297:2007	ISSN International centre http://www.issn.org/	+	+	+	+	eight digits ISSN 1330-7371
ISTC (International Standard Text Code)	textual work	2008	ISO 21047:2011	The International ISTC Agency http://www.istc-international.org	+	+	+	+	registration agency, year, textual work, check digit. 16 digits ISTC 0A9-2002-12B4A105-7
NBN (National Bibliography Number)	publications deposited and catalogued in national bibliographies which do not already have a publisher-assigned identifier	2001	NO	NO	+	+	+	+	country-specific NBN:fi-fe201003181510
SICI (Serial Item and Contribution Identifier)	parts of a serial	1991	ANSI/NISO Z39.56-1996	NO	+	+	+	+	Item, contribution, control segment variable length SICI 0015-6914(19960101)157:1<62:KTSW>2.0.TX;2-F
URN (Uniform Resource Name)	resource	1994	RFC 1737 RFC 2141	NO	+	+	+	-	namespace identifier, namespace specific string URN http://www.nsk.hr/

* Unique; *¹ Persistent; *² Resolution; *³ Metadata

DOI (Digital Object Identifier) is based on the international standard ISO 26324:2012 [3]. A DOI name can be used to identify objects of any material form (digital or physical) where these objects are content-related entities, as well as abstractions (such as textual works) [3]. DOI is made up of a prefix and suffix elements separated by a forward slash. There is no limit to the length of DOI. The prefix identifies the DOI registry and registrant, and the suffix can be a sequential number or can be based on another system used by the registrant. An example is *doi:10.1006/jmbi.1998.2354* for *Brian G. Turner, Michael F. Summers. "Structural Biology of HIV." Journal of Molecular Biology, 285(1), pp. 1–32. doi:10.1006/jmbi.1998.2354.*

ISBN (International Standard Book Number) is based on the international standard ISO 2108:2005 [4]. ISBN is a unique international identification system for each product form or edition of a monographic publication published or produced by a specific publisher [4]. ISBN consists of 13 digits comprising five elements: prefix element, registration group element, registrant element, publication element and check digit. ISBN 978-953-304-412-5 identifies a book (978) by a Croatian (953) publisher VBZ (304) by Krizmanić, Mirjana: *A što sad* (412, check digit 5) published in 2012.

ISSN (International Standard Serial Number) is based on the international standard ISO 3297:2007 [5]. ISSN is a unique identifier for a specific serial or other continuing resource in a defined medium [5]. Examples of serials include resources like journals, magazines, newspapers, updating web sites, and so on. ISSN is an eight digit number, including a check digit, for instance 1331-1182 identifies a journal *Croatian international relations review*.

ISTC [6] (The International Standard Text Code) is applicable to any textual work, whenever there is an intention to produce such a textual work in the form of one or more manifestations. It identifies content that can be published in many forms, languages, scripts, on a number of media, by numerous publishers anywhere in the world. ISTC consists of 16 numbers and/or letters with the following elements: registration element, year element, work element and check digit. ISTC is based on the international standard ISO 21047:2011. For example 42010111177788\$2 is ISTC for George Orwell's *Animal farm*.

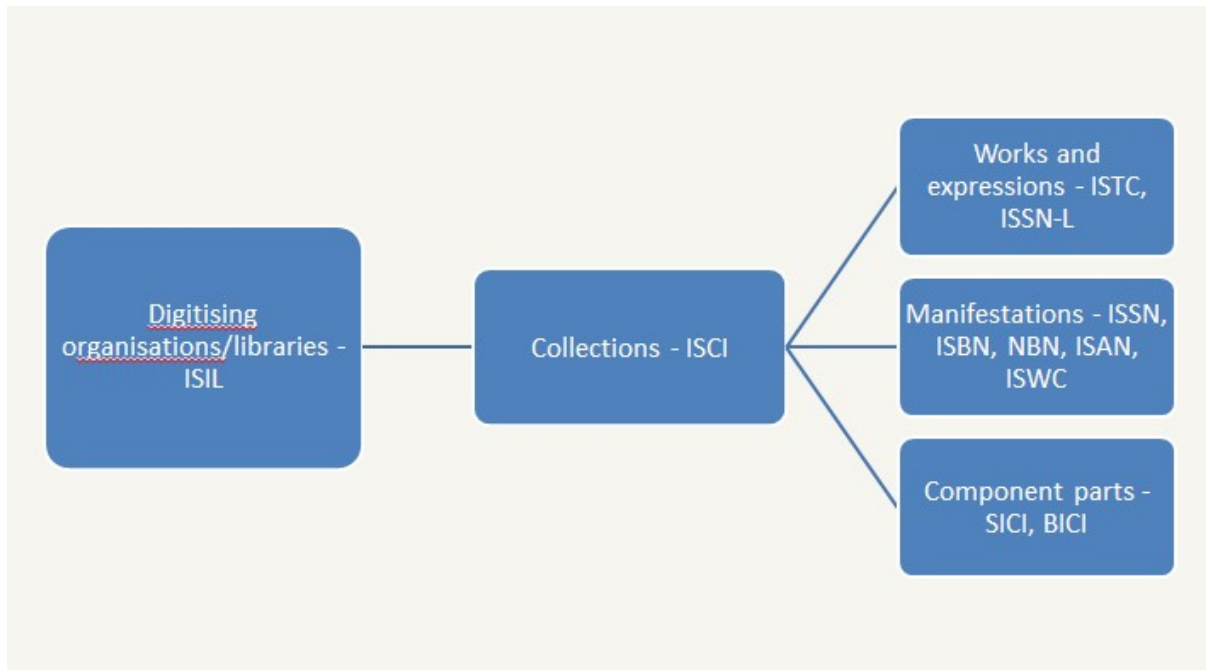
NBN [2] (National Bibliography Number) is a generic name for a group of publication identifier systems used by national libraries in some European countries (Finland, Germany, etc.). NBN is used to identify publications deposited and catalogued in national bibliographies which do not already have a publisher-assigned identifier. NBN consists of a combination of letters and digits. Each national library uses its own strings, and decides about the syntax and scope of NBN application independently of other national libraries and there is no international control or standard. NBN has a controlled prefix, a country code that makes each NBN unique. An example is NBN:SI:DOC-Q39YPQCQ for digitised issue of the newspaper *Clevelandska Amerika* Digital library of Slovenia (<http://www.dlib.si/?URN=URN:NBN:SI:DOC-Q39YPQCQ>).

URN (Uniform Resource Name) is intended to serve as a persistent, location-independent, resource identifier [7]. It is developed by IETF. The URN syntax is : <URN> ::= "urn:" <NID> ":" <NSS>. <NID> stands for the Namespace Identifier, and <NSS> for the Namespace Specific String. For example *urn:www.agxml.org:schemas:all:2:0* identifies *Agricultural Markup Language 2.0 for Grain and Oilseed Business*.

SICI [1] (Serial Item and Contribution Identifier) uniquely identifies parts of a serial publication (volume, article, etc.) in all formats. It is based on the ANSI/NISO Z39.561 (1996) standard. SICI has variable length and is composed of three segments (control, item and contribution segment). This is an example of SICI for an *abstract from Lynch, Clifford A. The Integrity of Digital Information; Mechanics and Definitional Issues. JASIS 45:10 (Dec. 1994) p. 737-44:*
SICI 0002-8231(199412)45:10<737:TIODIM>2.3.TX;2-M

4. How to identify digitised objects

Besides the imperative to identify digitised objects, there are no universal rules and the choice of identifiers will depend on the needs of users. As it was stressed earlier, identification is possible at different layers of granularity. This is illustrated by Picture 1.



Picture 1: Possible levels of identification

ISIL can be used to identify a library (or other heritage institution) that digitises. Each digitised collection can have an ISCI. Works (ISTC) and their individual manifestations (for instance ISBN for books or ISSN for serials) can be identified. Links between the work level and different manifestations are possible. Additionally, component parts of serials can be identified by SICI.

It must be stressed that the traditional identifiers ISSN and ISBN that were initially designed only for printed works have well adapted to digital environment as well. Each new edition of these ISO standards follows the development of resources and their different forms and manifestations. DOI and URN, however can be used on every level of identification and they can embed in their syntax any of the above listed identifiers.

This can be illustrated with the example of Jane Austen's famous novel *Pride and prejudice*. (see Picture 2)




Picture 2: Jane Austen: *Pride and prejudice*


The work level gets an ISTC. Different manifestations include different print editions, e-book on tablets or iPhones, audio book on a CD-ROM, but also digitised copies of early and old printed editions on different portals. Each manifestation gets its own ISBN, URN (for online editions) or NBN (in the unlikely case that no other identifier was assigned).

Identification of serials at different levels is somewhat different. The work level of the Croatian newspaper *Narodne novine* can be identified with ISSN-L, and each manifestation - print, CD-ROM, and online has its own ISSN. ISSN-L is an ISSN designated by the ISSN Network to group different media versions of the same serial. Online manifestations on the web, digitised editions and editions on mobile devices have one ISSN. The component parts, like a volume, issue or chapter can be identified by SICI, identifier for component parts of serials. (see Picture 3)

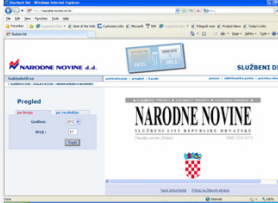
Work level: Narodne novine –
ISSN-L 0027-7932
Manifestation level: online, print, CD-ROM
editions



Narodne novine (Zagreb,
1876) ISSN 0027-7932



Narodne novine (Online)
ISSN 1333-9273



Narodne novine (CD-ROM)
ISSN 1331-7725

Picture 3: *Narodne novine*, work and manifestation levels

5. Conclusion

Precise and unique identification of digitised objects has become increasingly important. Identifiers can increase access, inform users, distinguish between different or similar objects, reduce duplication, link different manifestations, facilitate metadata transfer and links to other systems.

That is why it is necessary to identify digitised objects at as many levels as possible because this will facilitate the future use of these objects and management of digital libraries. More importantly, there is no need to invent new identifiers; the existing ones already fulfil all identification goals, and their further promotion and enhancement is recommended. This is especially true for traditional identifiers like ISBN or ISSN that have worked well for the last 40-50 years and have also proved their viability for digitised objects.

References

- [1] ANSI/NISO Z39.56 - 1996 (R2002) Serial Item and Contribution Identifier (SICI), [http://www.niso.org/apps/group_public/project/details.php?project_id=75_\(2012-08-28\)](http://www.niso.org/apps/group_public/project/details.php?project_id=75_(2012-08-28))
- [2] Hakala, Juha. Using National Bibliography Numbers as Uniform Resource Names, [http://tools.ietf.org/html/draft-hakala-rfc3188bis-nbn-urn-00_\(2012-08-28\)](http://tools.ietf.org/html/draft-hakala-rfc3188bis-nbn-urn-00_(2012-08-28))
- [3] ISO 26324:2012 Information and documentation -- Digital Object Identifier system (DOI)
- [4] ISO 2108:2005 Information and documentation -- International Standard Book Number (ISBN)
- [5] ISO 3297:2007 Information and documentation -- International Standard Serial Number (ISSN)
- [6] ISTC (The International Standard Text Code) home page, <http://www.istc-international.org/html/about.aspx> (2012-08-28)
- [7] URN Syntax, [http://www.ietf.org/rfc/rfc2141.txt_\(2012-08-28\)](http://www.ietf.org/rfc/rfc2141.txt_(2012-08-28))
- [8] Lynch, C., C. Preston, R. Daniel. Using Existing Bibliographic Identifiers as Uniform Resource Names, <http://www.ietf.org/rfc/rfc2288.txt> (2012-07-07)
- [9] Using The ISSN (International Serial Standard Number) as URN (Uniform Resource Names) within an ISSN-URN Namespace, [http://tools.ietf.org/html/rfc3044_\(2012-07-07\)](http://tools.ietf.org/html/rfc3044_(2012-07-07))
- [10] Wikipedia, [http://en.wikipedia.org/wiki/Identifier_\(2012-07-07\)](http://en.wikipedia.org/wiki/Identifier_(2012-07-07))

dgetliher@nsk.hr

jzajec@nsk.hr