



Identifiers: bridging language barriers

Jan Pisanski
Maja Žumer
University of Ljubljana
Ljubljana, Slovenia

and

Trond Aalberg
Norwegian University of Science and Technology
Trondheim, Norway

Meeting: **93. Cataloguing**

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY
10-15 August 2010, Gothenburg, Sweden
<http://www.ifla.org/en/ifla76>

Abstract:

Identification of bibliographic entities is an important part of bridging library world's linguistic and cultural barriers, as it allows for seamless use and reuse of bibliographic data in various applications. The paper looks at the existing potential candidates for identification of formal entities and discusses their suitability in terms of widespread use in libraries and beyond.

Introduction

As libraries deal with increasing quantity of information, it is important for users to clearly understand the environment they interact with. In terms of Functional Requirements for Bibliographic Records (FRBR), the new paradigm relating to the bibliographic universe, users need to be able to differentiate between instances of entities in order to perform their user tasks: finding, identifying, selecting, obtaining, as well as any other possible tasks not identified in FRBR, such as exploring. There is a growing need to uniquely identify the instances of all of the important entities in the bibliographic universe. As FRBR is the only formally recognized model of the bibliographic universe, it seems natural that entities that form the backbone of FRBR are to have identifiers. Of course, FRBR is a model that lends itself to different interpretations and is vague in some of its definitions. However, this does not imply that attempts should not be made to try to identify the different entities. While perfect identification is not always possible, due to the decidedly non-black-and-white nature of the bibliographic universe, having a system of distinct entity identifiers in place should help users find their way, no matter how one interprets the bibliographic universe.

Need for identification in libraries

Traditionally, libraries relied on combination of attributes as unique identification of library materials and other entities (e.g. headings for personal names, uniform headings). Although this works relatively well within the confines of a traditional library and within a limited environment, libraries can no longer afford redundancy in terms of duplicating cataloguing and authority control effort when libraries are isolated from each other. Additionally, in many cases, the application of identifying elements in individual catalogues has been poor, resulting in inconsistent data. Costs of labour are relatively high and the spreading use of information and communication technology provides a means for efficient and effective division of work, even on global scale. While local identifiers are normally good enough for local applications, the true strength of identification lies in internationally recognised identifiers. Users do not want to limit their searching to one bibliographic database and increasingly require integrated access to the broad range of resources available online, thus requiring harmonisation of bibliographic information in question. Finally, libraries need to be able to integrate and be a part of the emerging social web to be able to harvest and disseminate user contributed content as a complementary resource.

However, in such circumstances culturally dependent - including all language dependent and script dependent - identifiers are bound to have their drawbacks. Admittedly, users want, deserve and should be presented with culturally appropriate solutions when displaying information. The need to display information in a way that will help to identify it and the need to uniquely identify instances of entities are, however, two entirely separate issues. True unique identification is best done with language independent identifiers. Using names for global identification of any kind of entity is not a good solution. Not only can names change, which does not make them good candidates for unique identification, there is also the problem of names being culture and language dependent. While there may be some who feel that we can all use the same form of a name for identification, such a solution forces the will of one culture over all the others. It would be hard to find a solution that is acceptable for everybody. In fact, there is absolutely no need for identification to be done in this manner. As Library of Congress Working Group on the Future of Bibliographic Control (2008, p. 24) put it: "The use of language strings [...] as identifiers for both display and data manipulation hinders data exchange across languages and across different data communities."

Identifiers can form the basis of authority files and thus help eliminate redundancy in catalogues, as well as potentially make catalogues much easier to use. Whether authority files are actually helpful to the user depends greatly on the particular application. There are cases, where authority records are not connected to the bibliographic records. A solution, where a change in an author's name (e.g., as a result of marriage) leads to having to manually change all of the data in bibliographic records, should not be an acceptable one.

Additionally, identifiers figure to be even more important as new ways of using and reusing library-provided data arise. For instance, identifiers are the key to the successful integration of bibliographic data in Semantic Web and libraries increasingly interested in using Linked Data.

Although the need for identification in libraries is not a new one (Tillett, 2007), recently the digital age has brought many different initiatives, some of which are described in Vitiello (2004), Hakala (2006), Tillett (2007) and Babeu (2008).

Identifiers

Identification of entities is one of the most intriguing fields of application of identifiers in libraries and beyond. That IFLA Working Group on Functional Requirements and Numbering of Authority Records was charged with assessing feasibility of an International Standard Authority Data Number (ISADN) is a clear sign of recognition of importance of proper identification in library circles, spreading well beyond Group 2 entities. However, the group came to conclusion that establishing such a number is not feasible, although the report of the group itself (Tillett, 2008) does not contain detailed economic reasoning beyond the conclusion.

Three possibilities are discussed in the report. In addition to rejecting ISADN standard number as basically a fairly good, but costly, solution that is difficult to maintain, text-based identification (single authorized heading) is rightfully dismissed, and basing identification on clustering of authority files from various sources (akin to VIAF project) is discussed as a good option for future developments. Also, the report acknowledges the ongoing ISO's development of ISNI (International Standard Name Identifier) as a standard party identifier which could be of benefit to several communities.

It is important to gain knowledge whether existing identifiers could be used on a larger scale in libraries and are thus unfairly forgotten and whether new standardised identifiers are needed in view of FRBR and its application in library practice. Consistent use of identifiers, regardless of the form, but preferably as standardised as possible, should be the cornerstone of the somewhat unclear future of using born-FRBR and legacy data in parallel. Identifiers should also enable new ways of reusing bibliographic information and linking bibliographic information to other resources, facilitating the interoperability with other domains. However, one must not forget that cultural differences, which exist in interpretation of the bibliographic universe, may hinder the usefulness of widely used identifiers.

There are several existing international identifiers that aim to identify bibliographic materials. However, the focus of their authors was not on FRBR entities, but rather on trying to satisfy a variety of needs in a variety of communities. This can make it difficult to unequivocally establish what FRBR entity they try to identify, as can be seen from Table 1, which features assessments made by several prominent authors, although some of the discrepancy in Table 1 may be due to the assessments made prior to official recognition of some of these standards. As can be seen from Table 1, manifestation identifiers are the least problematic.

Table 1: Comparison of assessment of identification of FRBR group 1 entities

	VITIELLO (2004)	GATENBY (2008)	LEBOEUF (2005)	HAKALA (2006)
ISBN	M	M	M	M
ISSN	M	M	M	M
ISRC	E	M	E	
ISAN	W, E*	W	W	
ISWC	W	W	W, E	
ISTC		W	E	W, E
ISMN	M	M	M	
V-ISAN		M	E	

*Although it is not clear from the text, author may in fact be referring to ISAN identifying works and V-ISAN identifying expressions.

Manifestation identifiers are also by far the most widely used entity identifiers in libraries. Therefore it is sobering to realize that only 30% of WorldCat materials have an international identifier (Gatenby, 2008). If one considers WorldCat data to be a good representation of the state of current bibliographic records in the world, this number should be a cause of great concern, as it implies a certain lack of preparedness, even when it comes to manifestations.

Still, the relative success of ISBN and, on a lesser scale, other manifestation identifiers, such as ISMN, shows that identification can be done on a worldwide scale, if properly implemented. Its adoption has quietly helped everyone involved with books and discussions on the cost of the system that provides transparent identification are rare. However, for reasons of coverage (e.g., materials, age and geography), there remain a number of manifestations without an ISBN, so the solution definitely has its drawbacks, Holdsworth (2008) presents another problem: in FRBR terms, same ISBN is sometimes given to various manifestations of the same expression.

While identification of manifestations is relatively straightforward, identification of works and expressions is less clear, especially since the bibliographic universe covers all types of materials. Additionally, identifiers at these more abstract levels rarely appear on the manifestations themselves. Unfortunately, for libraries too often information that does not appear in or on the manifestation does not exist, even if it is in the best interest of the user. One must also not forget that so far all of the cataloguing has been done at the manifestation level and the need to identify works and expressions was less prominent. All of this makes identification of works and expressions almost non-existent in libraries at the present.

ISRC (International Standard Recording Code) is an expression identifier and ISAN (International Standard Audiovisual Number) is a work identifier, however they pertain to only a relatively small part of the bibliographic universe and are not particularly well used in libraries. On the other hand, nature of ISWC (International Standard Music Work Code) and ISTC (International Standard Text Code) is not as obvious. In ISWC arrangements, adaptations and translations receive their own ISWC (Antelman, 2004). In fact, ISWC Net database (www.iswc.org) shows highly inconsistent allocation of ISWCs, in terms of FRBR levels, but also in terms of geographical coverage.

On the other hand, the ISTC standard shows internal inconsistencies. As ISTC is applied to “textual works” and texts are associated with FRBR expressions, ISTCs should be allocated at this level. In fact, the examples in Annex E of the ISTC standard (ISO 21047, 2009) have different versions of the same textual work (e.g., revisions and translations in example E.4) allocated different ISTC numbers. From this we can gather that ISTC numbers are actually applied at the expression and not the work level in FRBR terms. On the other hand, this directly contradicts the statement in Annex B (B.13) of the published ISTC standard that the same textual work should not be allocated more than one ISTC. While the text of example E.4 describes relationships between a textual work and several textual works that are derived from it, the instances provided clearly indicate that translations and revisions are still considered to be the same textual work, whereas an adaptation for children of the “original” textual work is not.

Since we live in a digital world, where identification of digital and non-digital objects is equally important, we should not forget identifiers such as DOI. DOI (Digital Object Identifier) is a relatively well-used digital identifier for objects of intellectual property. It, however, does not presume identification of any particular entity. In fact, it can identify physical or digital manifestations, performances and abstract works (International DOI Foundation, 2006). From a FRBR-oriented point of view this means that even well-intended use of DOI may actually lead to greater confusion.

Conclusion

There are several different international standards for identification of the different parts of the bibliographic universe in existence. While some of them are fairly well used and applicable to a particular FRBR entity (e.g., ISBN, ISMN), most of the cogs for identification of FRBR entities on a global scale are not in place. Even more importantly, existing identifiers are meant to identify only fragments of the bibliographic universe (e.g., text, music, audiovisual materials), are underused, often even misused - although that may be due to the cultural differences - and are sometimes used to identify instances of different FRBR entities without a mechanism to properly identify the entities themselves. Additionally, in some cases it is difficult or costly for libraries to gain access to the existing identifiers. All of this makes efficient identification of FRBR entities using current identifiers in the near future highly unlikely. On the other hand, similar problems are likely to arise even with the adoption of any new identifiers.

The library community has grown to accept that internationally usable identifiers, ideally shared by different communities, are essential for a global integration of libraries. Which identifiers and what they will identify remains to be seen. While costs associated with maintaining a serviceable identification system are high and accurate identification may not always be possible, maintaining status quo may lead to even higher costs for libraries.

Bibliography

Antelman, K. (2004). Identifying the Serial Work as a Bibliographic Entity. *Library Resources & Technical Services*. 48 (4), 238-255.

Babeu, A. (2008). *Building a »FRBR-Inspired« Catalog: The Perseus Digital Library Experience*.

<http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf>

Gatenby, J. (2008). *The activities of OCLC on FRBR*. Workshop on FRBR in the European Library, 9. October 2008, Lisbon, Portugal.

http://frbr.bnportugal.pt/documentos/The_activities_of_OCLC_on_FRBR.ppt.

Hakala, J. (2006). The seven levels of identification. *Program*. 40 (4), 361-371.

Holdsworth, M. (2008). *The Identification of Digital Book Content*. Report prepared for the Book Industry Study Group, January 2008.

http://www.bisg.org/docs/DigitalIdentifiers_07Jan08.pdf

International DOI Foundation (2006). *DOI Handbook*.

<http://www.doi.org/hb.html>

ISO 21047 (2009). *Information and documentation – International Standard Text Code (ISTC)*, 22 p.

LeBoeuf, P. (2005). *Identifying 'textual works'*. FRBR in 21st Century Catalogues, Dublin, Ohio, May 2-4 2005.

<http://www.oclc.org/research/events/frbr-workshop/presentations/leboeuf/ISTC.ppt>

Library of Congress (2008). *On the record: Report of The Library of Congress Working Group on the Future of Bibliographic Control*.

<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>

Tillett, B. (2007). Numbers to Identify Entities (ISADNs – International Standard Authority Data Numbers). *Cataloging and Classification Quarterly*. 44 (3/4), 343-361.

Tillett, B. (2008). *A Review of the Feasibility of an International Standard Data Authority Number (ISADN)*. Prepared for the IFLA Working Group on Functional Requirements and Numbering of Authority Records, edited by G. Patton.

<http://archive.ifla.org/VII/d4/franar-numbering-paper.pdf>

Vitiello, G. (2004). Identifiers and identification systems. *D-Lib Magazine*. 10 (1).

<http://www.dlib.org/dlib/january04/vitiello/01vitiello.html>

International Standards Discussed in the Paper

ISO 2108 (2005). *Information and documentation – International Standard Book Number (ISBN)*, 21 p.

ISO 3297 (2007). *Information and documentation – International Standard Serial Number (ISSN)*, 20 p.

ISO 3901 (2001). *Information and documentation – International Standard Recording Code (ISRC)*, 9 p.

ISO 10957 (2009). *Information and documentation – International Standard Music Number (ISMN)*, 13 p.

ISO 15706-1 (2002). *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 1: Audiovisual work identifier*, 12 p.

ISO 15706-2 (2007). *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 2: Version Identifier*, 20 p.

ISO 15707 (2001). *Information and documentation – International Standard Music Work Code (ISWC)*, 10 p.

ISO 21047 (2009). *Information and documentation – International Standard Text Code (ISTC)*, 22 p.