



Identifier: Überbrückung von Sprachbarrieren

Jan Pisanski
Maja Žumer
University of Ljubljana
Ljubljana, Slowenien

and

Trond Aalberg
Norwegian University of Science and Technology
Trondheim, Norwegen

Translation:
Susanne Oehlschläger
Deutsche Nationalbibliothek

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:

Die Identifizierung bibliografischer Entitäten ist ein wichtiger Bestandteil zur Überbrückung von sprachlichen und kulturellen Barrieren in der Bibliothekswelt, weil sie eine nahtlose Nutzung und Nachnutzung von bibliografischen Daten in verschiedenen Anwendungen ermöglicht. Der Vortrag beschäftigt sich mit den vorhandenen potentiellen Kandidaten zur Identifizierung von formalen Entitäten und diskutiert ihre Eignung im Hinblick auf ihre breitgefächerte Anwendung in Bibliotheken und darüber hinaus.

Einleitung

Da Bibliotheken mit einer steigenden Informationsmenge zu tun haben, ist es für die Benutzer wichtig, die Umgebung, mit der Sie interagieren gut zu verstehen. Nach den Functional Requirements for Bibliographic Records (FRBR), dem neuen Paradigma des bibliografischen Universums, müssen Benutzer in der Lage sein, zwischen Instanzen von Entitäten zu unterscheiden, um ihre Anforderungen zu erfüllen: Finden, Identifizieren, Auswählen, Erhalten, und auch andere mögliche Anforderungen, die nicht in den FRBR identifiziert sind, wie zum Beispiel Erforschen.

Es gibt ein wachsendes Bedürfnis, die Instanzen aller wichtigen Entitäten im bibliografischen Universum eindeutig zu identifizieren. Da die FRBR das einzige formal anerkannte Modell

für das bibliografische Universum darstellen, scheint es naheliegend, dass Entitäten, die das Rückgrat der FRBR bilden, Identifier haben müssen. Natürlich sind die FRBR ein Modell, das sich für verschiedene Interpretationen eignet und in einigen seiner Definitionen vage ist. Allerdings bedeutet das nicht, dass keine Versuche unternommen werden sollten, die verschiedenen Entitäten zu identifizieren. Während eine perfekte Identifizierung nicht immer möglich ist, weil das bibliografische Universum eindeutig nicht nur schwarz und weiß ist, sollte die Tatsache, dass an dieser Stelle ein System von eindeutigen Entitäten-Identifiern zur Verfügung steht, den Benutzern helfen, ihren Weg zu finden, unabhängig davon, wie jemand das bibliografische Universum interpretiert.

Notwendigkeit zur Identifizierung in Bibliotheken

Traditionellerweise verlassen sich Bibliotheken auf eine Kombination von Merkmalen als einzige Identifizierung von Bibliotheksmaterialien und anderen Entitäten (d. h. Ansetzungen für Personennamen, Einheitsansetzungen). Obwohl das relativ gut funktioniert innerhalb der Grenzen einer traditionellen Bibliothek und innerhalb einer begrenzten Umgebung können Bibliotheken sich nicht länger Redundanz im Hinblick auf doppelte Arbeit für Katalogisierung und Normdatenkontrolle leisten, dadurch dass sie voneinander isoliert sind. Hinzu kommt, dass in vielen Fällen die Verwendung von identifizierenden Elementen in einzelnen Katalogen dürftig ist, was zu inkonsistenten Daten führt. Die Arbeitskosten sind relativ hoch, und die Ausbreitung des Einsatzes der Informations- und Kommunikationstechnologie liefert ein Mittel zur effizienten und effektiven Arbeitsteilung, auch weltweit.

Während lokale Identifier normalerweise gut genug sind für lokale Anwendungen, liegt die wahre Stärke für die Identifizierung in international anerkannten Identifiern. Benutzer wollen ihre Suchen nicht auf eine bibliografische Datenbank beschränken und fordern zunehmend einen integrierten Zugang zu dem breiten Spektrum der Ressourcen, die online erhältlich sind, und damit die Harmonisierung der in Frage kommenden bibliografischen Informationen. Schließlich müssen Bibliotheken in der Lage sein, sich zu integrieren und ein Teil des entstehenden sozialen Netzes zu sein, um die von den Nutzern hinzugefügten Inhalte zu harvesten und als kostenfreie Ressource zu verbreiten.

Allerdings haben unter solchen Bedingungen kultur-abhängige – einschließlich aller Sprachen- und Schrift-abhängigen – Identifier zwangsläufig ihre Nachteile. Zugegeben, die Benutzer wollen und verdienen kulturell geeignete Lösungen, wenn sie Informationen anzeigen lassen und diese sollen ihnen auch präsentiert werden. Die Notwendigkeit, Informationen auf eine Art und Weise anzeigen zu lassen, die dazu beiträgt sie zu identifizieren und die Notwendigkeit, Instanzen von Entitäten eindeutig zu identifizieren, sind jedoch zwei gänzlich getrennte Fragen. Echte, eindeutige Identifizierung geschieht am besten durch sprachen-unabhängige Identifier. Die Verwendung von Bezeichnungen zur globalen Identifizierung für irgendeine Entität ist keine gute Lösung. Nicht nur, dass Bezeichnungen sich ändern können, was sie nicht zu guten Kandidaten für eine eindeutige Identifizierung macht, es gibt auch das Problem, dass Bezeichnungen kultur- und sprachen-abhängig sind. Während es einige geben mag, die glauben, dass wir alle dieselbe Form einer Bezeichnung zur Identifizierung verwenden können, drängt eine solche Lösung den Willen einer Kultur über alle anderen auf. Es würde schwierig, eine Lösung zu finden, die für alle akzeptabel ist. Tatsächlich gibt es absolut keinen Grund dafür, die Identifizierung auf diese Art und Weise vorzunehmen. Wie die Library of Congress Working Group on the Future of Bibliographic Control (2008, S. 24) gesagt hat: „Die Verwendung von Textstrings [...] als Identifier sowohl

für die Anzeige als auch die Datenbearbeitung behindert den Datenaustausch über Sprachen und über unterschiedliche Daten-Communitys hinaus.“

Identifizierer können die Grundlage für Normdateien bilden und so dazu beitragen, Redundanzen in Katalogen zu reduzieren, ebenso wie möglicherweise Kataloge einfacher benutzbar zu machen. Ob Normdateien für Benutzer tatsächlich hilfreich sind, hängt im großen Maße von der jeweiligen Anwendung ab. Es gibt Fälle, in denen Normdatensätze nicht mit den bibliografischen Datensätzen verbunden sind. Eine Lösung, bei der eine Änderung bei einem Verfassernamen (z. B. nach einer Eheschließung) dazu führt, dass alle Daten in den bibliografischen Datensätzen manuell verändert werden müssen, sollte keine akzeptable Lösung sein.

Außerdem spielen Identifizierer eine sogar noch wichtigere Rolle, da sich neue Wege für die Nutzung und Nachnutzung von Daten eröffnen, die von Bibliotheken geliefert werden. Zum Beispiel sind Identifizierer der Schlüssel zur erfolgreichen Integration von bibliografischen Daten ins Semantic Web, und Bibliotheken sind zunehmend daran interessiert, Linked Data zu nutzen.

Obwohl die Notwendigkeit zur Identifizierung in Bibliotheken nicht neu ist (Tillett, 2007), hat das digitale Zeitalter jüngst viele verschiedene Initiativen hervorgebracht, von denen einige bei Vitiello (2004), Hakala (2006), Tillett (2007) und Babeu (2008) beschrieben sind.

Identifizierer

Die Identifizierung von Entitäten ist eines der faszinierendsten Anwendungsfelder von Identifizierern in Bibliotheken und darüber hinaus. Dass die IFLA Working Group on Functional Requirements and Numbering of Authority Records damit beauftragt wurde, die Machbarkeit einer Internationalen Standard-Normdatennummer (ISADN) zu untersuchen, ist ein deutliches Zeichen des Anerkennens der Bedeutung von richtiger Identifizierung in Bibliothekskreisen, die weit über die Entitäten der Gruppe 2 hinausgeht. Die Gruppe kam jedoch zu dem Schluss, dass die Einführung einer solchen Nummer nicht machbar ist, obwohl der Bericht der Gruppe selbst (Tillett, 2008) über die Schlussfolgerung hinaus keine detaillierten wirtschaftlichen Gründe enthält.

Drei Möglichkeiten werden in dem Bericht diskutiert. Zusätzlich zur Ablehnung der ISADN-Standardnummer als eine grundsätzlich recht gute, aber kostspielige Lösung, die zu schwierig zu verwalten ist, wird auch die textbasierte Identifizierung (einzelne Ansetzungsform) richtigerweise verworfen, und die Identifizierung auf der Grundlage von geclusterten Normdateien aus verschiedenen Quellen (ähnlich dem VIAF-Projekt) wird als eine gute Alternative für künftige Entwicklungen diskutiert. Auch anerkennt der Bericht die aktuelle Entwicklung des ISNI (Internationaler Standard-Bezeichnungs-Identifizierer) als ein Standard-Identifizierer an, der für einige Communitys nützlich sein könnte.

Es ist wichtig, Erkenntnisse darüber zu gewinnen, ob bestehende Identifizierer im großen Umfang in Bibliotheken verwendet werden können und deshalb zu Unrecht vergessen sind, und ob neue Standard-Identifizierer im Hinblick auf FRBR und ihre Anwendung in der Bibliothekspraxis benötigt werden. Die einheitliche Verwendung von Identifizierern unabhängig von ihrer Form, aber vorzugsweise so standardisiert wie möglich, sollte der Eckpfeiler der etwas unklaren Zukunft der parallelen Verwendung von echten FRBR- und bereits vorhandenen Daten sein.

Identifizier sollten auch neue Wege ermöglichen, bibliografische Informationen nachzunutzen und bibliografische Informationen mit anderen Ressourcen zu verknüpfen, indem sie die Interoperabilität mit anderen Ressourcen erleichtern. Allerdings darf man nicht vergessen, dass kulturelle Unterschiede, die bei der Interpretation des bibliografischen Universums bestehen, die Nützlichkeit von weit verbreiteten Identifiern einschränken können.

Es gibt einige bestehende internationale Identifizier, die darauf abzielen, bibliografische Materialien zu identifizieren. Allerdings hatten ihre Verfasser nicht die FRBR im Blick, sondern versuchten vielmehr eine Vielzahl von Bedürfnissen in einer Vielzahl von Communitys zufriedenzustellen.

Das kann es schwierig machen, eindeutig zu ermitteln, welche FRBR-Entität sie versuchen zu identifizieren, wie man aus Tabelle 1 sehen kann, die Einschätzungen darbietet, die von mehreren bekannten Verfassern gegeben wurden, obwohl einige der Unterschiede in Tabelle 1 den Einschätzungen geschuldet sein können, die vor der offiziellen Anerkennung einiger dieser Standards getätigt wurden.

Wie man aus der Tabelle 1 sehen kann, sind die Manifestation-Identifizier am wenigsten problematisch.

Tabelle 1: Vergleich der Einschätzung der Identifizierung von FRBR-Entitäten der Gruppe 1

	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	

*Obwohl es aus dem Text heraus nicht klar wird, kann der Verfasser sich tatsächlich darauf beziehen, dass ISAN Werke identifiziert und V-ISAN Expressionen.

Identifizier für Manifestationen sind auch bei weitem die verbreitetsten Identifizier für Entitäten in Bibliotheken. Deshalb ist es ernüchternd festzustellen, dass nur 30 % der Materialien im WorldCat einen internationalen Identifizier haben (Gatenby, 2008) Wenn man die Daten des WorldCat als ein gutes Abbild des Stands der bibliografischen Datensätze in der Welt betrachtet, sollte diese Zahl Besorgnis erregen, weil sie auf einen Mangel an Bereitschaft hinweist, selbst wenn es um Manifestationen geht.

Dennoch zeigt der relative Erfolg der ISBN und in kleinerem Ausmaß der anderer Identifizier für Manifestationen, wie dem ISMN, dass die Identifizierung weltweit erfolgen kann, wenn sie richtig implementiert ist. Ihre Einführung hat jedem, der mit Büchern zu tun hat, still und leise geholfen, und Auseinandersetzungen über die Kosten dieses Systems, das eine transparente Identifizierung bietet, sind selten. Aus Gründen des Geltungsbereichs (z. B. Materialien, Alter und Geografie) gibt es immer noch eine Reihe von Manifestationen ohne eine ISBN, so dass die Lösung definitiv ihre Nachteile hat. Holdsworth (2008) zeigt ein anderes Problem auf: nach FRBR ist dieselbe ISBN manchmal für verschiedene Manifestationen derselben Expression vergeben.

Während die Identifizierung von Manifestationen relativ einfach ist, ist die Identifizierung von Werken und Expressionen weniger eindeutig, insbesondere da das bibliografische Universum alle Arten von Materialien abdeckt. Hinzu kommt, dass Identifier auf diesen abstrakten Ebenen selten auf den Manifestationen auftauchen. Unglücklicherweise existieren für Bibliotheken allzu oft diejenigen Informationen nicht, die nicht in oder auf der Manifestation erscheinen, selbst wenn sie von großem Interesse für den Benutzer wären. Man darf auch nicht vergessen, dass bisher die gesamte Katalogisierung auf der Manifestationsebene erfolgte und die Notwendigkeit, Werke und Expressionen zu identifizieren weniger bedeutend war. All dies führt dazu, dass die Identifizierung von Werken und Expressionen gegenwärtig in Bibliotheken fast nicht existiert.

Der ISRC (International Standard Recording Code) ist ein Expression-Identifier und die ISAN (International Standard Audiovisual Number) ist ein Werk-Identifier, obwohl sie nur zu einem relativ kleinen Teil des bibliografischen Universums gehören und nicht besonders verbreitet sind in Bibliotheken. Auf der anderen Seite ist die Natur des ISWC (International Standard Music Work Code) und des ISTC (International Standard Text Code) nicht so offensichtlich. Beim ISWC erhalten Bearbeitungen, Adaptionen und Übersetzungen ihren eigenen ISWC (Antelman, 2004). Tatsächlich zeigt die ISWC Net Database (www.iswc.org) eine höchst inkonsistente Belegung mit ISWCs im Hinblick auf die FRBR-Ebenen, aber auch im Hinblick auf die geografische Abdeckung.

Andererseits zeigt der ISTC-Standard interne Inkonsistenzen. Weil der ISTC für textliche Werke vergeben wird und Texte mit den FRBR-Expressionen in Verbindung stehen, sollten ISTCs auf dieser Ebene vergeben werden. Tatsächlich aber enthalten die Beispiele im Anhang E des ISTC-Standards (ISO 21047, 2009) unterschiedliche Versionen desselben textlichen Werkes (z. B. Überarbeitungen und Übersetzungen in Beispiel E.4), denen verschiedene ISTC-Nummern zugeordnet sind. Daraus können wir schließen, dass die ISTC-Nummern tatsächlich im Hinblick auf die FRBR auf der Expressions- und nicht auf der Werk-Ebene zutreffen. Auf der anderen Seite widerspricht dies direkt der Aussage in Anhang B (B.13) des publizierten ISTC-Standards, dass dem selben textlichen Werk nicht mehr als ein ISTC zugeordnet werden sollte. Während der Text des Beispiels E.4 die Beziehungen zwischen einem textlichen Werk und mehreren textlichen Werken beschreibt, die davon abgeleitet sind, geben die dargebotenen Beispiele deutlich an, dass Übersetzungen und Überarbeitungen immer noch als dasselbe Werk angesehen werden, wohingegen eine Adaption für Kinder vom originalen, textlichen Werk nicht als solche angesehen wird.

Da wir in einer digitalen Welt leben, in der die Identifizierung von digitalen und nicht digitalen Objekten gleichermaßen wichtig ist, sollten wir nicht Identifier wie den DOI vergessen. Der DOI (Digital Object Identifier) ist ein relativ verbreiteter digitaler Identifier für Objekte geistigen Eigentums. Er maßt sich aber nicht die Identifizierung einer bestimmten Entität an. Tatsächlich kann er physische oder digitale Manifestationen, Aufführungen und abstrakte Werke identifizieren (International DOI Foundation, 2006). Aus einem FRBR-orientierten Blickwinkel bedeutet dies, dass selbst die wohlmeinende Anwendung des DOI tatsächlich zu einer noch größeren Verwirrung führen kann.

Schlussfolgerung

Es gibt mehrere unterschiedliche internationale Standards zur Identifizierung der unterschiedlichen Teile des bibliografischen Universums. Während einige von ihnen ziemlich

stark verbreitet sind und auf eine bestimmte FRBR-Entität anwendbar sind (z.B. ISBN, ISMN), sind die meisten Stellschrauben zur Identifizierung von FRBR-Entitäten global gesehen nicht an der richtigen Stelle. Noch wichtiger ist, dass die existierenden Identifier, die nur Fragmente des bibliografischen Universums identifizieren sollen (z. B. Text, Musik, audiovisuelle Materialien) zu wenig und oft sogar falsch genutzt werden – obwohl das kulturellen Unterschieden geschuldet sein mag – und manchmal dazu benutzt werden, Instanzen verschiedener FRBR-Entitäten zu identifizieren, ohne einen Mechanismus zu haben, um die Entitäten selbst richtig zu identifizieren. Dazu kommt, dass es in einigen Fällen schwierig oder teuer für die Bibliotheken ist, Zugang zu den bestehenden Identifiern zu erlangen. All dies macht die effiziente Identifizierung von FRBR-Entitäten durch die Verwendung bestehender Identifier in naher Zukunft höchst unwahrscheinlich. Auf der anderen Seite werden ähnlich Probleme wahrscheinlich auch bei der Übernahme irgendeines neuen Identifiers auftauchen.

Die Bibliothekscommunity akzeptiert zunehmend, dass international anwendbare Identifier, die idealerweise von verschiedenen Communitys geteilt werden, für die globale Integration von Bibliotheken essentiell sind. Welche Identifier und was sie identifizieren werden, bleibt zu sehen. Während die Kosten im Zusammenhang mit der Pflege eines funktionsfähigen Identifizierungssystems hoch sind und die genaue Identifizierung nicht immer möglich sein mag, kann die Beibehaltung des Status quo zu noch höheren Kosten für die Bibliotheken führen.

Literaturverzeichnis

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>

Internationale Standards, die im Vortrag erwähnt werden

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.