



标识符：跨越语言的障碍

Jan Pisanski, Maja Žumer
(斯洛文尼亚卢布尔雅那大学)

Trond Aalberg
(挪威科技大学, 挪威, 特隆赫姆)
中文翻译：陈宁 (中国国家图书馆)

Chinese Translator:
CHEN Ning
(National Library of China)

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>

摘要

书目实体 (bibliographic entities) 的标识是沟通图书馆界语言和文化障碍的一个重要组成部分, 因为这使得书目数据在各类应用中的无缝使用和再使用成为可能。本文分析了现有的几种标识正式实体的候选方案, 并对这些方案在图书馆界乃至更广泛领域的适用性进行了讨论。

引言

图书馆目前处理的信息量日益增长, 因此图书馆的用户需要清楚地了解他们所接触的这个环境。依据书目领域的新模型《书目记录的功能需求》(Functional Requirements for Bibliographic Records, FRBR), 用户需要能够区分不同的实体才能完成诸如查找、标识、选择、获取等任务, 以及其它在 FRBR 中没有指出的任务, 例如, 探索 (exploring)。在书目领域, 人们越来越多地需要对所有重要实体进行唯一标识。由于 FRBR 是书目领域内唯一被正式认可的模型, 构成 FRBR 核心内容的实体自然也就需要标识符。当然, FRBR 是一个可以有多种不同阐释的模型, 它自身的一些定义也是模糊的。然而, 这并不是说不应该去尝试标识不同的实体。我们并不是总能

做出完善的标识，这是因为书目领域在本质上并非泾渭分明。但尽管如此，不论人们如何阐释书目领域，建立一个清楚明晰的实体标识符体系可以帮助用户找到方向。

图书馆开展标识工作的必要性

图书馆过去一直依赖各种属性的组合作为标识图书馆资料和其它实体的唯一途径（例如个人名称标目、统一标目）。尽管这在传统图书馆以及在受限的环境下可以很好地发挥作用，但图书馆在相互隔绝的情况下已经无法再承受重复进行编目和规范控制工作的负担。此外，很多情况下，在独立目录体系中标识元素的应用并不尽如人意，造成了数据上的不一致。由于人力成本相对较高，广泛使用信息通信技术就为实现乃至全球范围内的有效分工协作提供了解决之道。尽管本地标识符（local identifier）对于本地应用来说通常效果不错，但标识体系的真正优势在于国际公认的标识符。用户不想把他们的检索局限在一个书目数据库，越来越多的情况下，他们要求对网络上的海量信息实现整合获取，这也就对书目信息的整合提出了要求。而且，图书馆需要整合新兴的社会网络并成之为其一部分，并收集和传播用户提供的信息作为补充资源。

然而，在这样的环境下，依赖文化因素的标识符——包括所有依赖语言和文字的标识符必然有其不足。必须承认，用户需要也应当获得在显示信息时能够恰当传达文化内涵的解决方案。然而，需要采用可以帮助标识信息的办法来显示信息，和需要对各种实体进行唯一标识是完全两回事。真正的唯一标识对于不涉及文化因素的标识符是最适用的。采用名称在全球范围内为任何一种实体作标识并不是一个好办法。名称会改变，对于唯一标识来说不是个好的选择；而且，名称涉及文化和语言内涵，也会带来问题。尽管有些人认为，我们可以都采用一个名称的同一种形式来进行标识，但是这种做法却是把一种文化的意愿强加给另一种文化。很难找到一个让所有人都接受的办法。事实上，也完全没有必要用这种方式来进行标识。正如“美国国会图书馆书目控制的未来”工作组所言（2008，P.24），“使用语言限制[...]作为显示信息和数据控制时的标识符阻碍了数据在语言之间和不同数据领域（data communities）之间的交换”。

标识符可以构成规范文档的基础，从而帮助消除目录中的多余信息，还有可能使目录变得更易使用。规范文档是否真的对用户有帮助很大程度上取决于具体的应用领域。的确存在这种情况，也就是规范记录和数据记录不关联的情况。由于责任者姓名的变化（例如，由于结婚）需要对书目记录中的所有数据进行手工修改的这种做法是不能接受的。

此外，由于出现了使用和重新使用图书馆所提供数据的新方法，标识符变得更加重要。例如，在书目数据成功整合到语义网的过程中标识符发挥了重要的作用；借助标识符，图书馆对于使用关联数据（Linked Data）也越来越有兴趣。

尽管在图书馆界使用标识符的需求并非新生事物（Tillett, 2007），然而新兴的数字时代却带来很多不同的方案，Vitiello（2004），Hakala（2006），Tillett（2007）和Babeu（2008）的文章中对这些方案中的部分内容有所涉及。

标识符

对于图书馆界乃至更广泛的领域而言，实体的标识是标识符最有意思的一个应用。国际图联规范记录的功能需求和编号工作组（IFLA Working Group on Functional Requirements and Numbering of Authority Records）受命评估国际标准规范数据号（ISADN）的可行性，这清楚地表明我们已经认识到在图书馆界进行恰当标识的重要性，这种重要性不仅局限于第二组实体，而是着眼于更广泛领域。然而，工作组得出的结论是，建立这种号码并不可行。不过，除了给出这个结论，工作组的报告（Tillett, 2008）并未从使用效益的角度做详细论证。

报告讨论了三种可能的方案。ISADN 标准号码未被采用，因为尽管这是一种相对较好的解决方案，但投入较大、难以维系；同时，以文本为基础的标识（单一规范标目）方案也被否决；和虚拟国际规范文档（VIAF, Virtual International Authority File）项目类似，采用基于汇集不同来源规范文档（clustering of authority files from various sources）的标识方案被认为是未来发展的一个不错选择。同样，报告也表示认同国际标准化组织正在进行的将国际标准名称标识符 ISNI（International Standard Name Identifier）作为标准责任者标识符的做法，因为这将使诸多领域受益。

非常重要的一点是，我们要搞清楚，现有的标识符是否可以在图书馆界更广泛的层面上使用，是不是这些现有的标识符被我们遗忘了；同时，也要搞清楚，就 FRBR 及其在图书馆领域的应用而言，是否需要新的标准化的标识符。对标识符的使用要保持一致原则，不论其形式如何，要尽可能标准化，这应成为未来同时使用原生 FRBR 数据和历史遗留数据的一个基础，尽管这一前景目前还不甚明朗。标识符同样还应该辅助实现新的方法，再使用书目信息并将书目信息与其它资源作关联，促进书目信息与其它领域的互操作性。然而，我们也不应忘记在书目信息阐释过程中存在的文化差异或许会影响被广泛使用的标识符的效用。

目前在国际上有几种用于标识书目资料的标识符。然而，这些标识符的作者所关注的重点不是 FRBR 实体，而是为了尝试满足不同领域的不同需求。这使得明确地对 FRBR 实体进行标识变得困难。正如表一所示，该表列出了几位著名作者所作的评估，其中也存在一些差异，这或许是因为评估是在这些标准得到官方承认前做出而造成的。从这张表我们可以看出，载体表现标识符（manifestation identifiers）是存在问题最少的。

表一：对 FRBR 第一组实体进行标识的评估结果作对比

	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	

* 尽管从文字上未作明确说明，但作者指的或许是用 ISAN 标识作品，用 V-ISAN 标识内容表达。

载体表现标识符也是目前在图书馆应用最广泛的实体标识符。因此，在 WorldCat 文献中只有 30% 具有国际标识符这一事实是发人深省的。如果说 WorldCat 数据可以较好地反映目前世界上书目记录的状况，那么 30% 这个数字应该引起我们的极大关注。因为这意味着缺乏准备，即使是对载体表现而言。

然而，ISBN 以及诸如 ISMN 这类应用范围更小的其它载体表现标识符的相对成功表明，如果实施恰当，标识是可以在全球范围内实现的。标识的采用潜移默化地帮助了所有与书籍打交道的人们，但是对于提供清晰明确的标识体系所需要的投入，相关的讨论却并不多见。由于覆盖面的原因（例如：文献类型、年代和地理分布），目前存在很多没有 ISBN 的载体表现，因此 ISBN 无疑也有其缺陷。Holdsworth (2008) 也提出了另一问题：在 FRBR 术语中，同一个 ISBN 有时候可以分配给同一内容表达的各种载体表现。

尽管载体表现的标识相对直接，但是作品和内容表达的标识就要模糊一些，特别是因为书目数据覆盖了所有类型的文献。此外，在这些更为抽象的层面上的标识符也很少在载体表现本身上出现。不幸的是，对于图书馆而言，在载体表现上没有出现的信息常常也就是不存在的，即使这个信息对用户来说是最需要的。我们不能忘记，目前为止，所有的编目工作都是在载体表现层面进行的，对作品和内容表达进行标识的需求并不突出。这也造成目前在图书馆界有关作品和内容表达的标识几乎没有。

ISRC (International Standard Recording Code, 国际标准音像制品编码) 是内容表达标识符，ISAN (International Standard Audiovisual Number, 国际标准音像作品编码) 是作品标识符。然而，这两种编码只和书目领域相对较小的一部分内容相关，在图书馆界并未得到很好的应用。另一方面，ISWC (International Standard Musical Work Code, 国际标准音乐作品编码) 和 ISTC (International Standard Text Code, 国际标准文本编码) 的性质并不明确。在 ISWC 的分配上，改编和翻译都获得独立的 ISWC (Antelman, 2004)。事实上，ISWC 网络数据库 (www.iswc.org) 在 ISWC 的分配方面存在很多相矛盾的地方，无论是在 FRBR 层面，还是在地理覆盖面上。

另一方面，ISTC 标准也体现出内在的不一致性。由于 ISTC 是应用于“文本作品”，而且文本又是与 FRBR 的内容表达相关，因此 ISTC 应该在作品层面进行分配。但是事实上，ISTC 标准附件 E 中的例子却对同一件文本作品的不同版本（例如，E.4 这个例子中的修订版和翻译版）分配了不同的 ISTC 号码。从这里我们可以看出，ISTC 号码实际工作中是被应用在了 FRBR 术语中的内容表达层面，而不是作品层面。同时，这也和已发布的 ISTC 标准中附件 B (B.13) 有关同一文本作品不能分配到多于一个 ISTC 的规定相矛盾。此外，E.4 这个例子也描述了一种文本作品和几种由其衍生的文本作品之间的关系，这些例子明确地表明，修订版和翻译版仍被认为是同一件文本作品，但是，面向儿童读者的对“原著”文本作品所做的改编版则不是。

我们生活在一个数字化的时代，数字化和非数字化对象的标识同样重要，我们因此不能忘记诸如 DOI 这样的标识符。DOI (数字对象标识符 Digital Object Identifier) 是一个相对得到较好应用的数字标识符，它所标识的对象都具有知识产权。然而，DOI 并不假定标识任何特定的实体。事实上，它可以标识物理或数字化的载体表现、表演

和抽象作品（国际 DOI 基金会，2006）。从面向 FRBR 的角度来看，这意味着即便是 DOI 的应用有着良好的初衷，它也会引起更大的混乱。

结论

目前有几种不同的国际标准用于标识书目领域不同方面的内容。其中有一些标准得到了较好的应用，也适用于特定的 FRBR 实体（如 ISBN，ISMN），然而主要用于在全球范围内标识 FRBR 实体的标准尚未制定。更突出的问题是，现有的标识符只用于标识书目领域很小的一部分内容（如，文本、音乐、视听资料），并且未被充分使用，或许是由于文化差异，经常还会被误用。有时，在没有恰当的机制对 FRBR 实体本身进行标识的情况下，现有的标识符还会被用来标识不同的 FRBR 实体。此外，在有些情况下，图书馆获取现有标识符也存在困难，或者说，费用较高。由于上述因素，在近期采用现有标识符对 FRBR 实体进行有效标识的做法是很难实现的。同时，采用任何新的标识符也会引起类似问题的出现。

图书馆界已经认识到，国际范围内适用并且可以跨界应用的标识符对于全球图书馆的整合至关重要。究竟会采用哪种标识符，这些标识符会用来标识什么，我们还要拭目以待。尽管维护一套可提供服务的标识体系费用不菲，而且也并非总能做到标识的准确无误，但是，维持现状、裹足不前只会使图书馆界付出更大的代价。

参考文献

- 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>

本文涉及的国际标准

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.