

FRBR 和 CIDOC CRM 知识本体之间的协调工作

□ 顾蓓 / 国家图书馆 北京 100081

摘要: 文章介绍了国际博物馆理事会下属的国际文献工作委员会提出的 CIDOC CRM 这个面向对象的概念参考模型以及最近成立的 FRBR 和 CIDOC CRM 国际协调工作组的工作。经过若干次会议, 工作组发现 FRBR 和 CIDOC CRM 都存在一些内在的缺陷。通过对双方的比较, 工作组提出了 FRBR 今后修订的一些建议。

关键词: 编目, FRBR, CIDOC, CRM, 知识本体, 概念模型

1 引言

目前的编目规则和编目实践基于“巴黎原则”和《国际标准书目著录》(ISBD)。尽管现在大多数图书馆都采用计算机编目, 但是其总体思路和 OPAC 显示还是从卡片目录发展而来的。一些研究人员认为, 用户目前在使用 OPAC 时候出现的许多问题, 都是因为 OPAC 目前仍然是卡片目录的电子版。

为了解决有关问题, 并适应多载体文献编目的需要, 国际图联在 1991 年至 1997 年间提出了《书目记录的功能要求》(FRBR) 这个实体-关系模型, 并于 1998 年由德国绍尔出版社 (K.G. Saur) 正式出版了这个文件。2002 年, 国际图联编目组 (IFLA Cataloguing Section) 任命了一个评估小组, 以维护这个模型, 并确保其进一步发展。FRBR 模型涉及“文字、音乐、地图、视听、图形、三维资料”, 其关键概念是分为三组的 10 个实体。第一组实体所代表的是图书馆目录中所描述的主要对象 (Object), 包括作品 (Work)、内容表达 (Expression)、载体表现 (Manifestation)、单件 (Item) 等; 第二组实体代表进行与第一组实体的事例 (Instance) 相关的活动的代理 (Agent), 包括个人 (Person) 和团体 (Corporate Body); 第三组实体代表作品所涉及的内容 (即与“作品”实体有“主题关系”的实体), 包括第一组实体和第二组实体中所涉及的所有实体, 还包括概念 (Concept)、对象 (Object)、事件 (Event)、地点 (Place)。FRBR 给

国际编目界带来了深远的影响,《国际编目原则》、ISBD、《英美编目规则》等编目领域内的主要文件都将在 2008 年左右完成修订, 采用 FRBR 的术语和思想。基于 FRBR 的目录将利用新的技术, 并成为一个全新的工具。

几乎与此同时, 国际博物馆理事会 (International Council of Museums, ICOM) 下属的国际文献工作委员会 (Comité International pour la Documentation, CIDOC) 也提出了 CRM (Conceptual Reference Model) 这个面向对象的概念参考模型, 提供文化遗产信息知识本体。它最初由 CIDOC 在 1996 年提出, 用以提供描述文化遗产文献工作中所用的概念和关系的定义和形式结构, 其目的是为异构信息库提供一种共同的语言, 以便在可能出现语义和结构不兼容的情况下进行数据集成。它目前由一个 CRM 特别兴趣小组 (CRM-SIG) 进行维护, 最新的版本是 2005 年 6 月发布的 4.2 版。它即将出版, 成为国际标准组织 (ISO) 的标准 (标准号 21127), 涉及博物馆在美术、考古学、自然史等领域内的所有类型的数据 (“著录”数据或“规范”数据), 也涉及博物馆、图书馆和档案馆所出版的各种文化遗产信息。

CIDOC CRM 是一个严格遵循知识本体原则的模型。由于有较长的历史, 所以发展得比较完善。此模型目前被多个本领域很有影响的项目所采用。自 2000 年起, 美国 RLG (研究图书馆集团) 开始考虑采用这个模型。但是, 中国图书馆界目前对其还不

十分了解。

2 CIDOC CRM

在FRBR模型中,图书馆编目中的“标目”指:概念、对象、事件、地点、个人、团体,而“书目记录”则描述了作品、内容表达、载体表现、单件。

在CIDOC CRM中,有60多个实体或类,它们都有各自的属性,属性表示了它们与其它实体之间的连接关系。这些实体组成了一个等级结构,其最上层的文化信息概念被映射到“物理实体”(Physical Entity)、“概念对象”(Conceptual Object)、“时间实体”(Temporal Entity)、“行动者”(Actor)、“地点”(Place)等类。此外,元数据的形式结构中还有如下一些概念:

- 称谓 (Appellations): 我们如何命名事物;
- 类型 (Types): 我们如何分类和组织事物形成受控的列表;

- 字符串 (Strings): 自由文本附注;
- 数字 (Numbers): 机器可处理;
- 时间元素 (Time Primitives): 机器可处理。

CIDOC CRM的中心概念是事件(Event)的概念:在空间和时间发生并且给世界带来某些变化的事物。

一个“事件”(即事件类中的一个事例)可以涉及如下方面:

- 行动者 (Actor) 类的事例,它们可能在引起事件的过程中起决定性的作用,或者只是目睹事件的进行,它们被称为“行动者称谓”(Actor Appellation)事例;

- 物理世界和/或心灵创造的碎片(即被称为“物理事物”(Physical Thing)这一类的事例和/或被称为“概念对象”(Conceptual Object)这一类的事例,例如画布和颜料以及画布和颜料形成的图像),它们被称为“称谓”(Appellation)事例(名称、头衔、代码等)。

此外,“事件”还有如下性质:

- 出现在时间中,所以有时间长度,即被称为“时间跨度”(Time-Span)这一类的事例,它们被称为“时间称谓”(Time Appellation)的事例(例如日期(Date)事例);

- 出现在空间中,所以与地点(Place)事例相关,被称为“地点称谓”(Place Appellation)的事例。

最后,还有一个CIDOC CRM中使用的一类“类型”(Type)概念,以将所有事物分类,并形成有关这个复杂、混沌而现实的世界的一个有组织的形象。

CIDOC CRM的创新之处在于:它不侧重围绕事物的描述,而是注重围绕事件的描述,将人、物质和非物质的事物在时空中连接起来;它清楚地描述了标识符和被标识事物之间的关系。但是它也有不足之处,例如缺少知识作品的模型。

在CIDOC CRM和FRBR之间,有一些相似的地方,但又不是——对应的。例如:

- “行动者”可以对应于“个人”和“团体”;
- “字符串”可以对应于“书目记录”里的附注;
- “对象”可以对应于“单件”;
- “称谓”可以对应于“标目”;
- “概念”可以对应于“作品”、“内容表达”和“载体表现”。

3 FRBR和CIDOC之间的协调

由于这两个模型都涉及图书馆和博物馆等领域,而且都试图成为标准,所以很有必要进行协调工作。2000年,在巴黎召开的欧洲图书馆自动化小组(ELAG—European Library Automation Group)会议上,图书馆界和博物馆界开始探讨协调FRBR和CIDOC CRM之间关系的可能性,当时Nicholas Criffts和Dan Matei提出了初步的FRBR模型的实体影射到CIDOC CRM类的面向对象的表示。经过若干年以后,图书馆界和博物馆界的代表于2003年组成了FRBR和CIDOC CRM国际协调工作组(International Working Group on FRBR/CIDOC CRM Harmonisation),该工作组的主席是Martin Doerr(由Patrick Le Boeuf协助),同时下属于国际图联编目组的FRBR评估组(FRBR Review Group)和CIDOC CRM特别兴趣组(CRM—SIG)。

研究的途径包括如下部分:

(1) 考察FRBR的每一个属性:

- 它的含义是什么?
- 关于这个含义有什么隐含的假设?
- 如果不是图书馆员,如何理解这个定义?
- 如何在类似CRM的结构中表达同样的含义?

(2) 图书馆员考虑的是什么?

● 编目过程有时候对于模型也很重要。例如编目员用大脑创建可控检索点的过程。

(3) 作者和出版社考虑的是什么?

● 要理解事物如何在那里发生。生产过程对模型十分重要。

(4) 如果一个实体有太多的属性, 就应该将其分解。

(5) 如果一个属性实际上就指一个事件, 那么就让事件更清晰一些。

(6) 编目员如何获得完全“抽象”实体的知识? (通过代表抽象实体的具体实体)

(7) 采用以用户(使用)为中心的途径。

协调工作组的主要目的是达成文化遗产信息在模型、标准、建议和实践等方面的共识。图书馆和博物馆都是“记忆机构”, 它们都致力于保护文化遗产对象和有关这些对象的信息, 它们有时候拥有同样的用户。它们之间的界限经常是十分模糊的: 图书馆拥有许多“博物馆对象”, 而博物馆也拥有许多“图书馆对象”。在这两种类型的机构中所保存的文化遗产对象产生于同样的文化背景或时期, 有时候由同样的代理创造。它们提供了可比文化特征(cultural features)的证据。所以, 很有必要建立针对这两种类型文化遗产机构所收集的信息的共同概念模型。

协调工具和语义网活动要求有一个集成的、共享的知识本体, 用于图书馆和博物馆的所有馆藏所积累的信息。这些馆藏是一个连续的整体, 从图书、CD、DVD等高度“标准化”的信息, 到植物或石头等“原始”材料, 还有手稿或雕版等“中间”物体。此外, 图书等“图书馆对象”可以是关于博物馆对象的, 而博物馆对象也可以表示图书中的事件或人物(例如博物馆藏的铭刻和青铜器等)。这样的关系或者被集成于共同的信息储存中, 或者至少以虚拟的形式通过中介手段集成起来, 允许查询语句同时不同的信息库里检索。这就需要有嵌入CIDOC CRM的FRBR_{oo}(面向对象的FRBR)这样的语义工具。

CIDOC CRM的修订过程受到了FRBR修订的影响, 考虑到了过去忽略的一些问题。同样FRBR的修订过程也受到了CIDOC CRM的影响。目前, 协调工作组开了八次会议(第八次会议于2006年10月召开), 但是还没有公布会议报告。工作组的工作主要是将

FRBR的实体和属性“翻译”成CIDOC CRM中使用的面向对象的形式, 但是有时候也反过来建议修订博物馆界提出的模型。不久以后, 将出现一个与CIDOC CRM一致的面向对象的FRBR模型的定义草案。

工作组有如下三个主要的原则:

(1) 采用以用户为中心的方法。

(2) 如果FRBR中的某个实体有过多的属性, 就说明该实体太复杂, 不能作为有关知识本体的原始概念, 这样的实体应该再被分解为更“原子”的概念。

(3) 如果一个属性实际上指一个事件(例如“日期”本身是没有意义的, 它必须是某个事件发生的日期), 就要努力使得这个事件更为明确。

例如, 在FRBR中定义的“作品”实体, 是“一种不同的知识或艺术创造”。“我们认识‘作品’, 是通过‘作品’的‘内容表达’来实现的, 但是‘作品’本身仅仅存在于‘作品’的各种‘内容表达’和‘载体表现’的内容的共性中”。这就导致“作品”的含义有若干可能的解释:

● 一系列个别“内容表达”所共有的概念;

● 一个出版物的总体概念(知识内容);

● 一个特定“内容表达”(=符号的集合)的知识内容, 不管其实现形式如何。

FRBR_{oo}(面向对象的FRBR)仍然保留“作品”这个一般的概念作为解释FRBR定义的超类(superclass)。F46“个别作品”(Individual Work)对应于与一整套符号相关的概念, F43“出版物作品”(Publication Work)即出版社对某出版物的知识贡献, F21“复杂作品”(Complex Work)更接近于FRBR中的主要解释。为了处理集成资源, 还定义了“容器作品”(Container Work)。因为这个方法更多地“基于事件”, 所以更侧重于创造过程。

FRBR_{oo}区别了两种不同类型的“内容表达”, 一种是表达其实现的作品之完全思想的“内容表达”, 另一种是表达部分思想的“内容表达”: 即F20“独立的内容表达”(Self-Contained Expression)事例和F23“内容表达片段”(Expression Fragment)。

例如, “载体表现”这个实体在FRBR中的定义是既可以解释为物理的事物, 也可以解释为概念的事物: 它既是“作品的内容表达的物理体现”, 也是“表示所有具有同样特征的物理对象”的实体; 它既是一个物理上的物品, 也是物理物品(一个集合)的精神

表示。“载体表现”既可以包括一件手稿（在这种情况下，“载体表现”就是“单件”），也可以包括一种出版物（在这种情况下，“载体表现”既是一个“类型”，也是一个“信息对象”（Information Object））。

FRBR_{oo} 努力解决这样的逻辑不一致问题，将“载体表现”这个实体拆分为两个不同的类别，对应于 FRBR 中“载体表现”模糊定义中的两种可能的解释，成为 F3“载体表现产品类型”（Manifestation Product Type）和 F4“载体表现独体”（Manifestation Singleton）。而 F3“载体表现产品类型”被作为 CIDOC CRM 类 E55“类型”的一个子类，从而也作为 CIDOC CRM 类 E28“概念对象”（Conceptual Object）这个仅仅是抽象的概念的一个子类；F4“载体表现独体”被作为 CIDOC CRM 类 E24“物理人造事物”（Physical Man-Made Thing）的一个子类，从而也作为 CIDOC CRM 类 E18“物理事物”（Physical Thing）的一个子类。

4 结论

协调 FRBR 和 CIDOC CRM 的好处是很多的。这项工作可以协调不同行业对文化遗产信息的观点，检查 FRBR 内在的一致性，促进可互操作性和集成性，促进 FRBR 和 CIDOC CRM 的共同发展。在以前的 FRBR 模型中，属性的定义是以现有工作为基础而提出的，可能会有问题。通过 FRBR 和 CRM 之间的比较分析，

人们确实发现有一些属性被分配给了错误的实体。

进一步的研究还涉及考察 FRBR 的所有其他实体（个人、团体、概念、地点、事件和对象）以及所有没有在 FRBR 中提及的 FRAR 实体，以后还要考虑 FR SAR。最后，形成一个正式的、稳定的 FRBR_{oo} 框架，提交 CIDOC CRM SIG 和国际图联 FRBR 评估组通过，再提交国际图联编目组通过。我们期望 FRBR_{oo} 将成为新的、“正式”的国际图联 FRBR 模型。

FRBR 于 2001 年被正式介绍到中国，中国图书馆界在 2004 年前后开始对 FRBR 模型投入了越来越多的关注，有一批有关 FRBR 的论文先后出现。但是，我们也应当正确认识 FRBR 的作用，既要看到它对国际编目界的发展所产生的深远影响，也不要忽视它内在的局限性，更不能盲目夸大其作用。FRBR_{oo} 是在西方文化的环境下形成的，有关研究人员很希望亚洲有人参与，使得这个模型的构想更全面一些。

从长远的观点来看，FRBR 模型还可能被扩展到表演艺术等更广阔的领域。

5 致谢

最后，本人要感谢法国国家图书馆的 Patrick Le Boeuf 先生。他亲自参加了有关的工作，并提供了有价值的资料。

参考文献

- [1] DOERR Martin, LE BŒUF Patrick (ed.). FRBR object-oriented definition (version 0.6.7); International Working Group on FRBR and CIDOC CRM Harmonisation supported by Delos NoE [M]. [S.l.]: [s.n.], August 2006.
- [2] DOERR Martin, LE BŒUF Patrick. Modelling Intellectual Processes: The FRBR - CRM Harmonization [C] // CIDOC 2006, Gothenburg, Sweden, September 12, 2006.
- [3] ŽUMER Maja, LE BŒUF Patrick. Conceptual models: museums & libraries: towards an object-oriented formulation of FRBR aligned on the CIDOC CRM ontology. [C] // ELAG 2006 "New tools and new library practices", Bucharest, 26 April 2006.
- [4] ŽUMER Maja, LE BŒUF Patrick. Harmonising without Harm: towards an object-oriented formulation of FRBR aligned on the CIDOC CRM ontology [C] // International Workshop on Semantic Interoperability for e-Research in the Sciences, Arts and Humanities, Imperial College Internet Institute, Imperial College, London, 30 March, 2006.
- [5] The CIDOC CRM [EB/OL]. [2006-11-01]. <http://cidoc.ics.forth.gr/>.
- [6] CIDOC CRM model [EB/OL]. [2006-11-01]. http://www.bnf.fr/pages/version_anglaise/normes/no-acCRM_gb.htm.
- [7] Tony Gill. Touring the RLG Information Landscape: the CIDOC Conceptual Reference Model [J/OL] // RLG Focus, August 2000 (45). [2006-11-01]. <http://www.rlg.org/legacy/r-focus/i45tour.html>.

- [8] 刘宏哲等. 基于 CIDOC CRM 的虚拟博物馆语义网络架构[J]. 计算机应用研究, 2006(4): 50-53.

作者简介

顾彬，管理学博士，研究馆员，国家图书馆图书采选编目部主任，国际图联编目组秘书兼常设委员会委员。通讯地址：中关村南大街33号国家图书馆图书采选编目部 100081

(收稿日期：2006-11-16)

Harmonization of FRBR and CIDOC CRM: Two Different Conceptual Models for Library and Museum Communities

Gu Ben / National Library of China, Beijing, 100081

Abstract: The author introduces CIDOC CRM, which is a conceptual model developed by CIDOC (Comité International pour la Documentation) of ICOM (International Council of Museums), and the work of International Working Group on FRBR/CIDOC CRM Harmonisation. After several conferences, the Working Group had some new findings about FRBR and CIDOC CRM, and proposed some recommendations for future revision of FRBR, which should be object-oriented.

Keywords: Cataloging, FRBR, CIDOC CRM, Ontology, Conceptual model