馆藏资源底层通用整体数据关联模型研究*

常　娥，华苏永

摘　要　细粒度、语义化和开放性是馆藏资源组织的三大需求。文章借鉴各种知识组织理论与方法，尤其是关联数据技术，尝试构建馆藏资源底层数据整体关联的通用架构。该模型由资源层、知识层和中间层等三层关联结构构成，提供更加一般化、普适性的资源组织框架，以充分满足馆藏资源、知识单元各自及其之间各种复杂的语义关系深度揭示与表达的需求。

关键词　知识组织　关联数据　通用模型

引用本文格式　常娥，华苏永. 馆藏资源底层通用整体数据关联模型研究[J]. 图书馆论坛，2016 (8)：7-12.

Bottom Common Organization Model of the Whole Library Knowledge Resource

CHANG E，HUA Su-yong

Abstract　Fine grit，semantization and openness play key roles in organizing library collection resources. The article constructs the bottom common organization model of the whole library knowledge resource on the basis of the various kinds of the knowledge methods，especially the linked data technology. The model is made up of 3 layers：links among the resources，links between resource and knowledge，and the links among knowledge. In the article，the common and universal resource organization frame is presented in order to express the complex semantic relation of resource and knowledge.

Keywords　knowledge organization；linked data；common model

0　引言

面对信息环境的变迁，图书馆馆藏资源正发生着深刻的变化，主要表现在3个方面：(1)馆藏资源发展迅猛、分布离散，成为集纸质文献资料、电子资源和网络资源于一体的综合馆藏资源体系；(2)馆藏文献资源基本完成由纸质文献到数字化资源的转换，并将不断朝着数据化的方向发展；(3)馆藏资源的类型不断拓展，不仅包括图书、期刊、报纸、手稿、研究数据等实际馆藏资源，还包括建立在实际馆藏资源之上的各种描述，即衍生馆藏资源，如书目数据、各种元数据方案、主题表、分类表以及规范文档[1]。

大数据时代，馆藏资源从形态到内容都发生了深刻变化，这给馆藏资源组织与利用带来重大挑战。图书馆试图从系统层面、元数据层面乃至资源层面将各种资源加以融合，然而无论是资源导航、OPAC整合、资源发现系统、网络信息聚

* 本文系国家社会科学基金项目“图书馆资源组织中的数据关联机制研究”（项目编号：14CTQ005）研究成果之一

合或信息共享空间等整合方式，均在数据融合范围、深度及效果等方面不尽如人意。关联数据(Linked Data)为突破馆藏资源组织困境提供了全新的思路，借助关联数据技术以细粒度、语义化、开放关联的方法组织各类馆藏资源，继而进行知识的深度挖掘、发现与利用，已成为图书情报学界研究的热点问题。然而关联数据的实际应用离不开知识本体的支持，因此目前图书馆界首先要解决的问题是尽快建立馆藏资源底层整体数据关联的通用框架，以实现网络环境下图书馆资源动态、关联、开放与多维的揭示与利用。

1　馆藏资源数据关联的理论基础

早在20世纪60年代，英国著名情报学家布鲁克斯就指出了文献组织与知识组织的区别，并认为理想的知识组织是对文献中的知识单元进行分析，找到人们创造与思考的相互影响与联系的结点[2]。但学者们同样敏锐地意识到，文献资源组织的实质就是一种知识组织[3]，文献资源组织和知识组织只是组织层次的区分而已[4]。文献资源组织和知识组织既有联系，又有区别。文献资源组织主要依据元数据标准对文献资源外部特征信息进行组织，而知识组织主要依据各种知识组织系统对文献资源内部知识进行组织。

根据布鲁克斯的设想，利用知识单元之间的逻辑关系表达，可将文献资源网络转变为由知识单元直接连接的概念网，这是一种理想的知识组织状态，是知识组织的最高目标。如果实现了文献资源的全语义化表达，在知识网络中即可进行各种知识单元的检索、推理和整合，那么知识网络是否可取代文献资源本身呢？这一问题值得我们思考。如果答案是肯定的，那么则无需在知识组织模型中建立知识单元与文献资源之间的关联，反之则需要建立这种联系，因此，这一问题的答案直接影响着馆藏资源知识组织模型的整体设计。

本文认为，无论知识网络中建立了多么丰富的关联，它始终无法取代文献资源本身，原因在于：知识即联系，对知识的组合与拆分形成了不同粒度的知识单元。文献本身就是一种知识单元，它是由人脑对细粒度知识单元进行分析、推理与综合等创造性活动后，得到的粗粒度知识单元。为了区分知识单元的粒度，一般使用文献代替粗粒度知识单元，使用知识单元代替细粒度知识单元。对于知识服务而言，何种粒度的知识单元最合适，目前学术界还未形成定论。从学术研究层面而言，文献的略读和精读一直都是交替进行的，计算机还无法取代人脑进行知识创新。因此，文献资源在过去、现在及未来都是人类最宝贵的知识资源。

王松林认为，目前文献资源仍是图书馆的组织对象[5]，这是图书情报学科与其他领域知识组织研究的重要区别。然而，馆藏资源组织不能仅停留在传统元数据方式的信息组织层面，而是要深入资源内容层面进行知识组织，这一点毋庸置疑。伴随着大量数字化科研成果的产出，馆藏资源不仅有图书、期刊、学术博客、研究数据等众多记录形式，而且往往以复合数字对象的形式存在[6]，需要同时处理文献资源、知识单元各自及其之间的各种复杂关联关系。

2　馆藏资源数据关联的技术基础

图书馆学领域一直致力于寻求文献、信息和知识序化与关联的最佳组织模型，从早期的目录索引，到MARC机读目录、DC元数据标准、本体模型(Ontology)，再到关联数据。以MARC机读目录为代表的现行元数据标准，是一种以文献为基本单位的粗粒度资源组织方法，因无法解决资源描述的异构性和语义性问题[7]，所以无法使图书馆摆脱信息孤岛的束缚。而本体模型虽可实现图书馆资源的语义化描述和知识组织，但在资源关联的广度和深度上未能提供更多的帮助。此外，元数据方法和本体模型，均不支持资源的开放获取。

近年来，关联数据技术的迅猛发展，以及图书馆界新一代编目标准RDA的发展，为馆藏资源最佳组织模型的探寻指明了新的发展方向。然而关联数据仅仅是一个技术框架，在此基础之上

还需融入元数据和本体模型，以解决资源描述的异构性和语义性问题，进行知识与资源的深度关联。美国、德国、法国、瑞典等国家图书馆纷纷发布了包含书目、规范主题词、规范人名等资源在内的关联数据集[8]。由于这些数据集以特定馆藏资源为对象，缺乏整体的数据关联框架，因此数据集关联的范围和程度有限，有些仅是内部关联，与非图书馆数据集的粘合度并不高，难以成为开放数据网络中的交通枢纽或核心节点。

以 FRBR 模型为核心的新一代编目标准RDA，正处于实施与应用的初级阶段[9]，虽设计了关联准备，然而如何利用关联数据技术为 RDA 的元素、概念词表及其相互关系提供表达、描述和管理——是选择将完整的 RDA 编目数据发布为关联数据，抑或选择其核心元素进行转换，仍需进一步研究[10]。

国内在理论研究方面主要是国外关联数据的概念、技术框架、典型数据集等相关研究的引介以及我国图书馆利用关联数据重组馆藏资源开展知识服务的思考[11-12]。学者们对于关联数据的发布与消费技术、开放应用协议等进行了分析[13-15]，并探讨了基于关联数据的馆藏数字资源语义聚合模型[16-17]。在应用实践方面，虽然国内尚未有图书馆彻底将书目资源发布为关联数据集[18]，关联数据中枢 Datahub 中收录的中文关联数据集极少，但上海图书馆和中国科学技术信息研究所等机构搭建了关联数据实验系统，其中上海图书馆在 2015 年底将家谱数据发布为关联数据，并于 2016 年 3 月首家率先推出提供了标准消费接口的家谱关联数据开放平台，以提供家谱数据的开发与利用[19]，这标志着我国图书馆界对于关联数据的研究已经进入应用阶段。

总体来说，关联数据为馆藏资源细粒度、语义关联的开放组织奠定了技术基础，国内对于关联数据的应用实践刚刚起步，还处于借鉴国外先进经验阶段。

3 底层通用的整体数据关联模型构建

馆藏资源的组织不能仅停留在传统元数据方式的信息组织层面，而是要深入到资源的知识内容层面。从文献资源的整体揭示与组织，深化到文献中的数据、公式、事实、结论等知识单元的关联组织，这是科技进步与发展带来的必然结果。然而，在知识控制单位重心发生转变的过程中，并不是非此即彼的，需用全面的视角来研究知识单元与文献资源间的组织与关联，即需揭示和表达文献资源、知识单元各自及其之间的三重复杂语义关联，以满足知识整体且快速的联通，以及检索与发现的功能。

因此，本文在综合借鉴各种知识组织理论与方法，充分考虑与现有相关元数据标准的兼容性，在 FRBR/FRAD/FRSAD 三者集成与扩展的综合概念框架之上，通过引入包含研究背景、材料方法、模型假设、实验数据、结果讨论等项目的学术元数据框架，构建图书馆资源底层通用的整体数据关联模型(Bottom Common Organization Model of the Whole Library Knowledge Resource，以下简称“ BCOM 模型”)，详见图 1。BCOM 模型为馆藏资源、知识单元各自及其之间的各种复杂语义关系建立了 3 层关联结构，即资源层、知识层和中间层，极大地增强了馆藏资源数据关联网络的整体连通性，无论是从资源出发，还是从知识单元出来，都能迅速关联并发现用户所需的知识或资源。

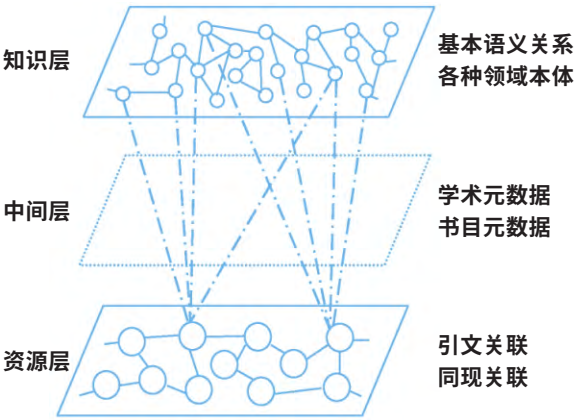

图 1 底层通用的整体数据关联模型

3.1 资源关联层

该层主要揭示资源与资源之间的关联关系。在 DC、MARC 等传统资源组织模型中，资源之

间的关联关系是粗糙的、隐含的。例如，DC 模型通过“相关”属性建立了资源之间的关联，然而没有具体定义资源间语义关系的类型。在BCOM模型中，将资源关联作为独立层次进行构建，以增进资源之间关系的语义表达与快速连通。

BCOM 模型将采用 RDF 框架和 URI 标识符描述资源之间的关系，即以“资源 - 属性 - 值”三元组形式进行描述，并对第一组元“资源”赋予 URI 号码。由于该层描述的是资源与资源之间的关系，因此第三组元“属性值”的取值是资源，同样应赋予 URI 号码。资源关联层主要通过文献之间的引用与概念同现建立关联。

3.2 资源与知识关联层

该层又称为中间层，主要揭示资源与知识单元之间的关联关系，在同现关联的基础上，附加学术元数据和书目元数据，建立各种语义关联。

在中间层，BCOM 模型致力于将资源的内容特征描述深入化，同时兼顾外部特征描述，从而形成了学术元数据和书目元数据这两套不同的元数据标准。其中书目元数据主要借鉴和吸收了 MARC 和 DC 相关元数据项，用以描述资源外部特征，而学术元数据则将重点探寻比关键词方式要具体，比文献资源本身要凝练的一种内容表达框架。换言之是将 MARC 和 DC 中的关键词和分类号等主题描述元素项抽取出来，进行具体化，以充分揭示资源的主题内容。

在中间层，BCOM 模型仍将采用 RDF 框架和 URI 标识符描述资源与知识之间的关联关系，其中第二组元“属性”即为学术元数据项和书目元数据项，第三组元“属性值”的取值是知识单元。知识单元是否应设置 URI 标识符的问题将在下一小节中讨论。

3.3 知识关联层

该层主要揭示了知识单元之间的关联关系。无论是复杂的知识组织系统(知识本体)，还是简单的知识组织系统(词表)，语义关系主要包含等级关系、等同关系、相关关系和并列关系这 4 种类型。不同学科领域对这 4 种类型语义关系进行更加细致、具体而深入地表达，从而构建了各种领域知识本体。

在知识层，BCOM 模型将融合 FRAD 与 FRSAD 模型中的人、机构、概念、实物、事件、地点等概念，以及 FRBR 模型中的作品层概念，重新梳理、定义各知识概念间的基本语义关系。所谓基本语义关系是指与领域无关的事物之间通用的语义关系，包括人物关系、地点关系、时间关系，以及通用主题概念关系等。此外，BCOM 模型将采用开放的方式，逐步构建领域知识本体，以进一步丰富知识层概念之间的语义关系。

值得一提的是，科学研究数据是一个特例，它既可归属于资源层，又可归属于知识层。本研究将从文献(著作、论文)中分解出的研究数据作为知识单元，利用学术元数据框架与资源层文献进行关联和映射，而将围绕某项科学研究，实验前后所产生的一系列、整套数据作为一种资源，除了在资源层利用引用和同现方式建立各种关联外，还将利用书目元数据框架与知识层概念建立联结。

在知识层，BCOM 模型仍将采用 RDF 框架和 URI 标识符描述知识间的关联关系，其中第二组元“属性”除了包含基本语义关系外，还将包含各种领域概念关系，第一组元“资源”和第三组元“属性值”的取值都是知识单元。由于该层包含人、机构、实物、地点、时间、事件等各种知识概念，对其进行规范控制十分必要，因此 BCOM 模型对存在同义现象的知识单元将赋予 URI 符号进行标识，以解决语义异构问题。

4 BCOM 与 RDA、BIBFRAME 框架的对比分析

1998 年国际图联构建 FRBR 本体模型，定义书目描述的 4 层结构，即“作品(work)- 内容表达(expression)- 载体表现(manifestation)- 单件(item)”。RDA 是 FRBR 模型的实践者，因此 RDA 中的元素设计全部与 FRBR 模型的层次结构相对应。然而有机构和学者始终认为 RDA 过于复杂，对其推广与应用前景十分担忧。2011 年，美国国会图书馆(LC)设计并推出了新的书目

框架计划 BIBFRAME(Bibliographic Framework Initiative)，它将 FRBR 模型中的 4 层结构简化为“作品(work)- 实例(instance)”两层结构，并且增加规范数据(authority)和注释数据(annotation)以适应规范控制和扩展描述的需求[20]。

显然，BIBFRAME 是作为 RDA 的竞争者出现的，BIBFRAME 对 RDA 的概念层次进行了简化，与 RDA 不完全一致。然而这种简化其实是一种更一般化，使得 BIBFRAME 成为一个更大的容器，对其进行扩展，完全可以容纳 FRBR 模型所有的描述功能。同时，BIBFRAME 还吸收了 MARC/MODS、DC、Schema.org 中的元素定义，共设计了约 300 个属性，分属 52 个类。

然而 BIBFRAME 的开发尚未完成，还存在许多不确定性，主要是规范控制和注释模型的争议较大。2015 年 6 月，在 BIBFRAME 词表修订中，LC 曾建议取消规范控制类及其属性，将规范原有子类直接归属于顶级类资源，如将 ISBN 等均作为顶级类[21]。由此可见，国际图联、美国国会图书馆，以及 OCLC 对于开放数据时代，以关联数据为基础的馆藏资源描述与组织框架仍在不断探索之中，谁也无法预言 RDA 和 BIBFRAME 最终发展形态。

无论是 RDA，还是 BIBFRAME，都将资源与知识单元放在同一个层次进行讨论，未深入揭示资源的主题内容，并且对于主题概念间的关系描述较弱，仍侧重书目关系描述。BCOM 模型则将资源与知识单元分层处理，单独构建了资源层、知识层关联，并将资源的内容特征描述和外部特征描述进行分离，形成中间层元数据标准。这不仅强化了资源主题内容的揭示功能和主题概念关系的描述功能，而且增强了整个资源组织框架在资源层和知识层的整体连通性。这是 BCOM 模型相比于 RDA 和 BIBFRAME 模型的重要区别。

此外，除了增强馆藏资源数据网络的整体连通性外，BCOM 模型还具有一定的领域通用性，主要体现在以下 3 个方面：(1)资源层、知识层和中间层，这 3 层架构对于任何领域的资源组织都是通用的，利用 BCOM 模型无论是从资源，还是从知识单元出来，都能迅速关联并发现用户所需知识或文献资源；(2)学术元数据框架和书目元数据框架也是领域通用的，旨在连通资源层与知识层；(3)采用 RDF 和 URI 描述资源、知识概念及其之间各种语义关系，满足了通用、语义化、开放与互联的需求。

最后，值得一提的是，由于 RDA 和 BIBFRAME 框架侧重资源外部书目属性表达，因此其语义关系是相对固定的，而 BCOM 模型则将具有学科差异性，因为要想完全发挥 BCOM 模型资源组织与发现的功能，需强化知识层概念语义关系的表达。BCOM 模型在知识层将采用完全开放的形式，构建各种领域知识本体，以丰富知识层各种语义关系的表达。未来本研究将借鉴 BIBFRAME 和 DCMI 应用纲要机制，制定 BCOM 模型的领域纲要，以适应特定领域知识本体的构建需要。纲要将规定领域知识如何继承和分解，可以复用哪些词表和属性等，使 BCOM 成为一个更大的通用“容器”。

5 结语

现代研究表明，大脑神经细胞的数量约有 150 亿之多，它们之间形成了极其复杂的网络结构，彼此沟通，相互影响，每个细胞与其他细胞可产生 2000 多种联系[22]。科学知识存在普遍联系，序化并还原知识间的连接关系是知识组织的重要目标。相比于人类神经细胞间的广泛联系，在传统人工组织的知识系统中，每个知识节点与其他节点的联系却十分有限，远远无法满足知识普遍联系、高速传递的需要。因此，需要对粗粒度的知识进行碎片化处理与挖掘，提取出研究对象、背景、问题、目标、方法、过程、结果、结论、讨论；分析出概念、原理、观点、规范、标准、设计、技巧、事实、数据、人物等知识单元，并揭示文献资源、知识单元各自及其之间的各种复杂语义关联，从而构建复杂知识关联网络，以使文献资源及其所承载的知识在开放数据网络中，得到广泛而深入的利用。有鉴于此，综

合借鉴各种知识组织理论与方法，尤其是关联数据技术，本文尝试构建馆藏资源底层整体数据关联的通用架构，即通用数据关联模型 BCOM，该模型旨在提供更加一般化、普适性的资源组织框架，以充分满足馆藏资源、知识单元各自及其之间各种复杂语义关系的深刻揭示与表达需求。BCOM 模型通过丰富知识节点间的联系，使人工知识网络尽可能接近真实、客观存在的知识网络，以促进知识广泛而深入的利用。

参考文献

[1] 常娥，孟祥保. 图书馆资源组织中的数据关联特征研究[J]. 图书馆论坛，2016 (2)：49-56.

[2] 文庭孝，刘晓英，刘进军. 知识关联的理论基础研究[J]. 图书馆，2010 (6)：9-11.

[3] 王子舟，王碧滢. 知识的基本组分——文献单元和知识单元[J]. 中国图书馆学报，2003 (1)：5-11.

[4][5]王松林. 图书馆组织对象及其层次研究[J]. 中国图书馆学报，2010 (1)：40-44.

[6] 白海燕. 基于关联数据的信息组织深度序化初探[EB/OL]. [2016-03-23]. http：//www.caigou.com.cn/lib/ziliao_detail.asp？id=20201.

[7] 刘炜，李大铃，夏翠娟. 元数据与知识本体[J]. 图书馆杂志，2004 (6)：50-54.

[8] Library Linked Data [EB/OL]. [2016-03-23]. http：//datahub.io/group/about/lld.

[9] OCLC 的 RDA 政策声明[EB/OL]. [2016-03-23]. http：//catwizard.net/posts/20120229210233.html.

[10] Byrne G，Goddard L. The strongest link：Libraries and Linked Data[J]. D-Lib Magazine，2010 (9).

[11] 黄永文. 关联数据在图书馆中的应用研究综述[J]. 现代图书情报技术，2010 (6)：1-7.

[12] 刘炜. 关联数据：概念、技术及应用展望[J]. 大学图书馆学报，2011 (2)：5-11.

[13] 夏翠娟，刘炜，赵亮，等. 关联数据的发布技术及其实现——以 Drupal 为例[J]. 中国图书馆学报，2012 (1)：49-57.

[14] 夏翠娟，刘炜. 关联数据的消费技术及其实现[J]. 大学图书馆学报，2013 (3)：29-37.

[15] 张春景，刘炜，夏翠娟，等. 关联数据开放应用协议[J].中国图书馆学报，2012 (1)：43-48.

[16] 欧石燕. 面向关联数据的语义数字图书馆资源描述与组织框架设计与实现[J].中国图书馆学报，2012 (1)：58-71.

[17] 游毅，成全. 试论基于关联数据的馆藏资源聚合模式[J]. 情报理论与实践，2013 (1)：109-114.

[18] 钟远薪，李田章，刘炜. OPAC 混搭关联数据应用研究[J]. 现代图书情报技术，2013 (6)：25-29.

[19] 黄启哲. 上图推出“家谱知识服务平台”[N]. 文汇报，2016-02-18 (9).

[20] 苏建华. 浅议 RDA 与 BIBFRAME[J]. 现代情报，2015 (1)：156-158.

[21] BIBFRAME 打算取消“规范”核心类[EB/OL]. [2016-03-20]. http：//catwizard.net/posts/tag/bibframe.

[22] 林格. 教育，要敢于面对事实[N]. 教育时报，2009-02-07 (4).

作者简介 常娥，东南大学图书馆副研究馆员；华苏永，东南大学图书馆馆员。

收稿日期 2016-04-06

责任编辑：付伟棠