DOI: 10.13530/j.cnki.jlis.150014

万维网时代的规范控制*

刘 炜 张春景 夏翠娟

摘 要 图书馆书目控制的理想是“搜罗并整序人类所有知识”，“规范控制”是从图书馆编目工作中发展起来的一项工作。然而规范控制从来没有很好地实现设计时的初衷，根本原因是思想太超前，而技术不成熟，又低估了人为执行规则并保证其一致性的难度。当前的语义万维网技术为规范控制提供一种绝好的实现平台，关联数据技术提供了概念与其表示完全独立的表达模型，可以基于书目信息中的所有属性特征，进行规范的、基于概念的检索。许多大型机构都在利用新的技术开发和研究规范控制服务，例如美国国会图书馆的 BIBFRAME 书目框架格式草案、OCLC 的虚拟国际规范档，以及 Open Authority 项目和 Wikidata 项目等。这些努力反映了一种发展趋势，图书馆行业数百年积累起来的书目控制经验，如果能充分利用好现代信息技术所提供的强大工具，不仅能实现过去没有实现的理想，而且能在更大范围内发扬光大。图 4。表 1。参考文献 22。

关键词 规范控制 书目控制 语义万维网 关联数据

分类号 G254

Authority Control for the Web

LIU Wei, ZHANG Chunjing & XIA Cuijuan

ABSTRACT

“Organizing and accessing all kinds of knowledge of mankind” is the ultimate goal for bibliographic control. Authority control as the major means for bibliographic control, evolved with library cataloguing at the very beginning. The collection of authority records called authority file is one of MARC standard formats. Authority files play a vital role in the traditional library catalog system. It refers to the preferred value of headings, and determines the precision ratio of retrieval, and provides an authoritative reference for the future cataloging. In order to implement authority control, there must be two elements: a set of cataloguing rules and a set of authority records which are very essential. So the traditional authority control contains a set of rules and records to regulate the title, creator and keywords and so on.

Many international organizations such as IFLA and national libraries have made great efforts in biblio-

* 本文系上海市哲学社会科学课题基金项目“基于书目控制的网络信息资源的规范控制方法”(编号: 2009BTQ001)和国家社科基金项目“关联数据的理论和应用研究”(编号: 11BTQ041)的研究成果之一。(This article is an outcome of the project “The authority control of network information resource based on bibliographic control” (No. 2009BTQ001) supported by philosophy and social science foundation of Shanghai and the project “Study on the theory and application of linked data” (No. 11BTQ041) supported by National Social Science Foundation of China.)

通信作者: 刘炜, Email: wliu@libnet.sh.cn, ORCID: 0000-0003-2663-7539 (Correspondence should be addressed to LIU Wei, Email: wliu@libnet.sh.cn, ORCID: 0000-0003-2663-7539)

graphic control for decades and invested enormous labors. A lot of large authority files including millions of names and vocabularies have been published, which play a significant role in organizing recorded knowledge. However, authority control has never accomplished the original goal of "Organizing all the knowledge for the mankind". There have been different nations with different rules, and a lot of major attributes without authority controlled values. The defect of manual work and the imperfection of catalogue rules are the main enemies. But the fundamental cause of the flaws is that the technology at that time can not provide the adequate support for the purpose of authority control. The difficulties to achieve a strict consistency in implementing the cataloging rules have been greatly underestimated.

The nature of authority control is to keep the consistency of bibliographic records. The "description" including two procedures: the first is to understand, abstract and conceptualize the properties featured in bibliographic objects; the second is to represent the properties with semiotic system such as a kind of written language. So the nature of authority control is to construct a relation between the conceptualization and the representation. Bibliographic system should provide a authority control mechanism to strictly conform to the consistency of concept representation, and make it independent of people, time or other environmental factors.

The emerging semantic web technology has just provided an excellent platform for authority control. The Linked Data technology offers an abstract model to separate the concept from its representation. It provides a capability to search all the attribute fields in bibliographic records as controlled values. Authority files can be encoded with RDF, SKOS or OWL schema. Thus, all cataloguing rules and specific rules (RDA/AACR2) can be encoded with semantic statements and integrated into the process of authority control.

Many international organizations are implementing and providing authority control services with semantic technologies, for example, BIBFRAME of Library of Congress, VIAF project of OCLC, Open Authority project, Wikidata project, etc.. These projects reflect a trend that the knowledge organization experiences of traditional libraries are still valuable for the information management in the internet era. The merits of bibliographic control in librarianship have contributed a lot to our society in history, and it will be still enlightening our civilization with the new information technology and foster our career to the utmost. 4 figs. 1 tab. 22 refs.

KEY WORDS

Authority control. Bibliographic control. Semantic web. Linked data.

1 传统规范控制的困境

规范控制是因图书馆编目而发展起来的一项工作，是图书馆“书目控制”思想的具体实践和重要内容，有时也与“书目控制”概念混用，旨在保障书目系统中重要属性描述的一致性，满足准确查找、辨识、归类和判定的要求。规范控制所涉及的书目属性通常有：人名、机构名、会议名、连续出版物名、图书题名以及主题词等[1]。图书馆在长期的编目实践中对规范控制逐渐形成了一整套规则和做法，并规定了用来提供规范控制依据的特殊的 MARC 数据，称为规范记录，大量的规范记录汇集在一起，称为规范档。高质量的规范控制通过对同名异形进行归一，对异名同形进行区分，并对词间关系进行导引指示，为高质量的书目系统提供了必要的保障。规范控制可以说是图书馆学对知识组织最为独特的贡献。

然而，长期以来，图书馆界对规范控制的认

识基本上是工具、现象层面的，这种认识上的缺陷，正在成为规范控制利用最新信息技术，并应用到更广泛相关领域的障碍。从业界对规范控制的定义中我们就可以感受到这种局限，例如以下定义。

规范控制是指图书馆编目或书目记录中所使用的保持标目（个人或团体名称、文献或丛编题名和主题）一致性的程序。该程序将规范文档应用于新增文献并将其加入馆藏[2]。

权威控制是通过使用权威词表（称为权威档），对图书馆目录中或书目记录文件中标目的一致性（包括名称、统一题名、连续出版物名和主题）进行维护，应用于新记录加入馆藏的过程[3]。

规范控制是为确保标目在检索款目及书目系统中的唯一性和稳定性而建立、维护、使用规范款目和规范文档的工作过程[4]。

这些定义是图书馆界非常典型的认识，所描述的都是为达到检索一致性的目的，在一定历史条件和技术条件下，通过长期实践总结出来的方法和手段，其中还沿用了 MARC 所代表的磁带数据文件时代甚至卡片目录时代的许多概念。这样虽然也确实解释了规范控制，但与具体的技术紧密结合起来，用具体做法代替了定义，没有说明为什么，没有抽象出规范控制的实质。

如果仅仅根据上述定义来认识规范控制，则规范控制就深深打上了 MARC 时代的烙印，越来越陷入人工处理高成本的泥潭中无法自拔。另外，在规范控制的效率、一致性效果以及影响面和影响范围等方面不仅无法突围，甚至难达初衷。尤其是，规范控制无法利用技术进步带来的全面网络化、智能化、社群化的优势，更无法应用到图书情报之外的广阔天地。传统的规范控制不能随当前信息技术进步取得相应进展，是规范控制面临的最大困境。

那么，规范控制的本质是什么？如何在各种技术条件下（尤其是万维网环境下）达到一定程度的一致性？规范控制的功能需求是什么？万维网环境下是否还有可能进行规范控制？是否应该有一个“度”？这是本文将要探讨的内容。

2 规范控制的本质

规范控制的本质是实现基于概念的描述和匹配。解释这个问题首先要从传统目录发挥功能的机制说起。

图书馆目录是馆藏的缩影和指代物，是读者与馆藏资源之间的桥梁。图书馆对馆藏的有序组织主要体现在其目录体系中。传统图书馆把卡片目录的功能发挥到极致，创造性地采用“标目”方式（即将该属性置于端首）组织目录体系（排序），这样一套馆藏可以有多套目录与其对应，从而提供了针对不同属性（著者、题名、主题、分类、机构、会议名、期刊名等）的不同检索途径，只需将这个属性作为标目即可。清账造册是大多数涉及仓储管理的行业都有的技术，然而只有图书馆的卡片目录，以上述这种方式，突破了财产清单或查检式目录在排序、互见、多对多对应方面的局限，非常灵活，这成为图书馆行业独有的创造。

所有的书目记录都包含三个要素：标目（即检索点）、书目描述和位置信息。检索点引导读者找到其所需要的书目记录，书目描述信息让读者判断是否是其所需要的资源，如果需要则位置信息提供了获取方式。这是编目工作所需满足的基本功能需求。

由此可见，“标目”是编目工作中最为重要的一项工作，直接关系到能否充分、准确揭示馆藏，能否建立起不同馆藏属性标目之间的关联关系等，也即关系到目录系统的质量，是一项“技术活”。规范控制即是对标目的一致性进行规范工作的总称，主要有两方面的工作。①规范记录（规范档）的编制和维护工作。其中要详细记录异名同义、同形异义或概念名称之间的关系等需要规范的信息，最好能与编目系统以及读者查询系统建立连接，才能更好地发挥规范数据在编目和查询方面的功能。②编目人员应用规范档确定正确标目形式的工作。原则上，理想状态下，这样做之后，应该能够在图书

馆的目录体系中，或读者检索时，将同一个作者的作品归并，将同一个作品的不同版本归并，将同名的不同作者的作品区分开，将同一主题及其上下位主题归并在一起，等等，其他属性也以此类推，从而实现规范控制的“汇集、区分、导引”的基本功能。

情报检索理论中有一个基本假设，就是任何语词都是概念的表征。当人们看到狗这种动物时，在大脑中就建立起狗这个概念，然后通过“狗”这一文字符号进行表征，这时“狗”就成了概念的文字标签，即规范词，表示的是概念本身，而不再是一个自由词。这样就用规范的语词或符号构造了一个概念空间，在其中所有的检索都可以认为是概念检索，即知识检索。

规范控制实际上就是这个理论的一个应用。通过编目人员所编制的规范档（记载了概念与概念表达——即语词或符号——之间的关系）来建立规范的概念空间：相同的概念有相同的表达，不同的概念有不同的表达，关系密切的概念应该能够用一定的表达明确地描述出它们的相关关系。传统的规范控制方法希望通过一整套规则、方法和规范档，建立一种人为的规范控制机制，应用于编目和检索系统中。如图1所示，当编目人员建立了人名规范档之后，书目系统就会自动将鲁迅、茅盾、巴金与他们的本名和其他众多笔名联系起来，这样就能使读者更准确、全面地检索到想要的文献。

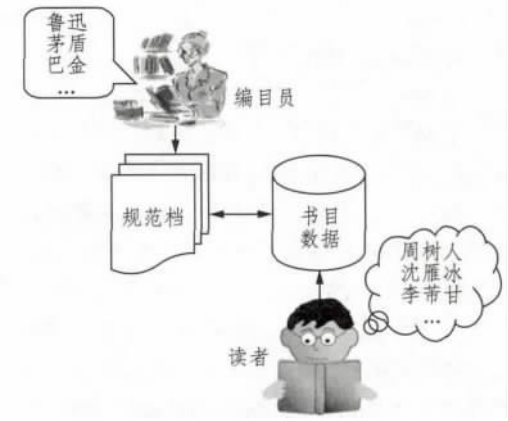

图1 规范控制的作用机制

3 传统的规范控制为什么不可能成功？

图书馆几乎自诞生之日起就以“搜罗并整序人类所有知识”为己任，这可以追溯到亚历山大图书馆。图书馆人经过长期的努力发现，可以通过编制全球统一书目而达到掌握人类所有知识的目的。虽然只是目录，但未必不能观照某一历史时期的全球知识，这是一个可行的权宜之计。这就是图书馆书目控制的理想。1950年，联合国教科文组织和美国国会图书馆对书目控制的定义是：从书目的目的出发，控制人类已出版的全部文献[5]。

从各类编目规则对于编目对象“文献类型”（GMD）的定义可以看出这个雄心：在以印刷资料为主要知识载体的工业化社会，图书馆的编目对象几乎是所有的知识产品。GMD包括的内容如表1所示。

表1 ISBD推荐的GMD[6]

文献类型	英文GMD标识	中文GMD标识
专著出版物	printed text	印刷文本
古籍（善本）	printed text	印刷文本
连续出版物	printed text	印刷文本
测绘制图资料	cartographic material	测绘资料
乐谱	printed music	印刷乐谱
电子资源	electronic resource	电子资源
图卡	graphic	图卡
全息照片	hologram	全息照片
缩微制品	microform	缩微制品
电影制品	motion picture	电影制品
模型制品	object	模型制品
录音制品	sound recording	录音制品
录像制品	videorecording	录像制品
投影制品	visual projection	投影制品
配套资料	kit	配套资料
多载体	multimedia	多载体

书目控制有两个层次的基本职能：了解世界上总共有多少知识产品；了解某个图书馆具体有哪些馆藏，以及如何获得这些馆藏。前者是宏观上的需求，图书馆的国际性联盟组织（如国际图联）长期致力于此，通过各种“书目控制”的行业规定和技术手段力图实现这个目的；后者是微观上的要求，各国国家图书馆或地区、行业、专业性的大型图书馆，对本国、本地区或本领域的图书馆负有责任，这是图书馆保存性职能的体现，也是图书馆职业的基本要求。要实现书目控制，前提是要有统一的编目规则和数据格式（卡片也可以作为一种记录格式），同时要有一套操作规程，规范控制就是实现书目控制的必要手段和方法的总称，有时也被当作书目控制的同义词。要实现规范控制，规则和规范档是两个必不可少的要素。长期以来，各级各类图书机构对书目控制进行了大量的研究和实践，尝试了很多技术，制订和维护了大量的原则标准和规则规范。

国际图联等各类国际组织和各国的国家图书馆数十年来在规范控制方面做出很大努力，投入了巨大的人力成本，数百万条规范名称和大量规范词表对于书目信息的有序组织起到了重要作用，使得经过图书馆编目的数据明显比其他来源的数据更具可信度。

然而规范控制的美好理想，从来没有不折不扣地实现过，人工进行规范控制的这种业务模式在信息社会已显得不合时宜。这主要表现在世界范围内规范控制的标准远未统一且各国参差不齐，规范控制的标目字段并未实现全覆盖，规范数据的数量极其有限，质量差强人意，等等，规范控制应用的一致性程度和深度都没有达到人们所期望的规模和水平。随着信息爆炸和知识载体类型的复杂化，规范控制的成本越来越高，效果得不到体现，性价比不高，全面实现规范控制的可能性越来越渺茫[7]。以至于美国国会图书馆在2007年末发布的《书目控制未来报告》中几乎宣布要放弃承担书目控制的责任[8]。

相对于图书馆的宏大梦想，图书馆所能支配和掌握的资源及武器少得可怜。这种反差一直伴随图书馆成百上千年的历史，直到今天。起源于卡片目录时代的规范控制思想，大大超前于卡片目录所能提供的技术手段，因为建立规范档以及将其应用于编目和检索的复杂过程，大都需要人工完成，而人是最不可靠的。由人来制订规则、执行流程、操控机器，不仅效率极低、成本巨大，而且根本无法保证一致性。在计算机技术突飞猛进发展的时代，规范控制方法还一味地模仿卡片目录时代的做法，也成为其最大的桎梏。

可以总结的原因还有很多，如人们在当时的认识存在局限，编目原则和编目规则本身有瑕疵，各国由于语言和文化的差异造成编目实践的特异，以及执行中存在的大量细节无法详细规定等。从总体上看，真正的原因其实只有一个：思想太超前，而技术尚不成熟，又低估了人为执行规则并保证其一致性的难度。正是这个原因造成规范控制在MARC时代成为“不可能的任务”。卡片目录时代，图书馆对于知识组织的领先优势，已经成为其后来跟不上技术进步潮流的包袱和阻力。

总之，书目控制理论所提出的目标过于庞大和理想，以卡片目录的管理为核心思想而发展起来的一整套信息描述和知识组织技术以及工作流程，远不足以支撑规范控制理想的乌托邦。

4 新技术带来新希望

根据摩尔定律，我们知道近半个多世纪以来，成本不变的情况下，计算机芯片的集成度一直呈指数增长，带来计算速度、存储能力和网络带宽的飞速发展，使我们正来到一个万物互联、“智慧”无所不在的崭新时代的入口。这个时代带给图书馆的，将是重新审视如何完成历史赋予的各项职能。例如，当知识以脱离载体的多媒体形态四处游荡时，图书馆该如何捕获、处理、保存、组织和提供它，并传之后代？

如同产业技术革命解放了人类的体力一样，信息技术革命正在极大地拓展人类的脑力。计算机首先解放了人脑的记忆功能，其次让人的交流不再只依靠文字这种经过抽象的媒介，还可以通过视频、音频甚至触觉、味觉（通过各类传感器）等更加本源的方式进行。得益于各类手持设备、可穿戴电脑和物联网，将来以视频等原始信息进行交流的方式将越来越大行其道。对图书馆而言，最大的变化是计算机已不仅仅能够实现将图书先扫描为图片，图片识别成文字，再对文字进行处理的模式，那样只能进行字符匹配，实现全文检索，而且可以以语义标注的方式直接对“知识”进行编码，这样就能进行知识处理和检索了。这就是说计算机已经能接管以往只有人类在大脑中才能建立的概念空间，开始以知识为编码和处理对象，并辅之以逻辑计算，进而向真正的智慧化挺进了。

现在看起来上述预测似乎还很遥远，但孕育着这一切的技术有些已经蛰伏了近 20 年，对于互联网技术而言，20 年已经跨代了。目前万维网（World Wide Web）作为互联网技术最成功的应用，已成为人类从事信息活动的垄断性平台，移动互联网也是其基本协议的延伸应用，所有的新技术、新应用、新模式都必须“触网”才可持续并被最多的人群了解。这其中包括以 RDF 和知识本体为代表的语义万维网技术，以社会性网络、群众智能为代表的 Web 2.0 技术，以及大数据、云计算、商业智能技术等。这些技术经过学术界和产业界的不断打磨，在软硬件和网络环境的合谋之下，现在终于到达了一个爆发临界点。在这种背景下，前述规范控制所面临的技术短板可望得到根本性的克服，MARC 时代的不可能任务有望变成可能。

综上所述，传统编目工作中的规范控制过程可抽象为从符号体系到概念体系的映射过程（见图 2）。书目系统的建立过程就是应用概念体系规范符号体系的过程，规范档的建立则是把符号体系抽象为概念体系的过程[3]，而读者的检索过程就是在后台用概念体系匹配符号体系，然后把匹配结果提供给读者的过程。只有将这一系列过程充分去除人工因素，实现流程化、自动化，才能保证高效、低成本和准确性，才能使规范控制可持续并得到拓展。万维网协会（W3C）十多年来不断完善语义万维网架构和众多的标准规范，尤其是用以表达语义的 RDF 模式和它们的扩展（如 SKOS、OWL 等），以及以 RDF 数据模型为基础的“关联数据”技术等，其目的正是构建概念化的知识空间，这与规范控制

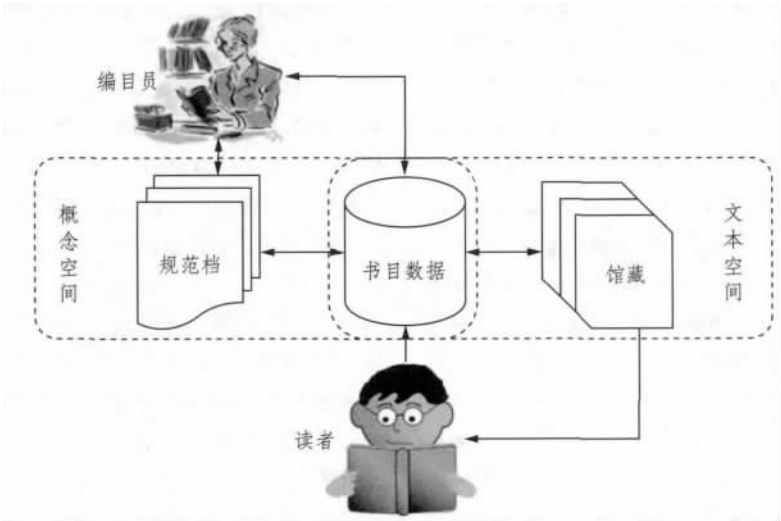

图 2 规范控制模型示意

的功能需求有着完美的契合，如果我们能结合当前日渐成熟的社会性网络，充分利用众包和群众智慧提供的信息自我完善机制，并把整个知识生产的流程纳入元数据语义获取和模型化的过程中来，规范控制的前景将一片光明。

5 关联数据如何满足规范控制的功能需求？

从20世纪90年代开始，国际图联为顺应书目控制应用环境的变化，对图书馆书目数据的功能需求进行了重新审视，采用计算机领域流行的实体—属性分析方法，提出了一个颠覆以往的概念模型，并先后推出了三个功能需求报告，分别是书目记录的功能需求（FRBR）、规范数据的功能需求（FRAD）和主题规范数据的功能需求（FRSAD），后两个报告直接针对规范控制。

实体—关系分析是构造模型的常用方法。计算机模拟现实世界必须首先建立模型，对同一事物，基于不同目的，可以用不同的观察角度和方法得到不同的模型，也就能解释不同的现象和因果关系。国际图联的这三个功能需求报告，都是围绕着第一个FRBR报告中提出的书目数据模型展开的，该模型将与书目数据相关的实体分为三类，详细分析了这些实体的相关属性和关系。这三类实体分别是：第一类书目实体，有作品、内容表达、载体表现和单件（WEMI）四种类型，第二类代理（agent）实体，有个人、家庭和团体三种，第三类为主题实体，包括概念、实物、事件、地点四个方面。报告提出书目记录的功能需求（用户任务）为查找（find）、辨识（identify）、选择（select）和获取（obtain），规范数据（authority data）的功能需求是查找（find）、辨识（identify）、提供情境（contextualize）和证明（justify），而主题规范记录（subject authority data）的功能需求是查找（find）、辨识（identify）、选择（select）和探索（explore）[10]。

国际图联的系列报告把图书馆书目控制带入了一个新时代，从此人们看待目录的方式与以往有了很大不同。FRBR在属性揭示的基础上纳入立体化的实体—关系分析，厘清了许多书目属性的隶属关系，更接近真实世界，同时更易于采用最新的计算机及网络技术进行模拟。这些功能需求报告所提出的理论很快被业界接受和采用，体现在许多标准规范中，如作为编目规则的RDA和作为MARC数据格式替代者的BIBFRAME（书目框架）等。虽然，能否借此成功将图书馆书目数据带入网络世界尚未可知，但已经有了很多有益的尝试，OCLC已经将其WorldCat书目数据FRBR化，并开发了支持关联数据服务的VIAF规范档系统[11]，美国国会图书馆也宣布将停用MARC并启用BIBFRAME，并已把大量的规范词表以关联数据形式发布到网上[12]。

不论上述系列报告中所提出的具体功能需求是否精当，或者是否还应该包括更多的需求，其满足需求的程度直接取决于规范控制的质量，具体来说，即检索点选择、名称控制、规范记录的丰富性以及参照引用是否充分及方便快捷等，而这些方面关联数据技术提供了天然的解决方案，主要表现在四个方面。

（1）关联数据技术提供了概念独立于其表示形式的表达模型。可以URI标识概念，以标签或名称属性表示各种语言或符号的表达，从而使“标目”问题得到完美解决，即不需要选择任何一种优先形式（如鲁迅）作为标目，标目就是代表概念的URI，任何同义词符号都可以作为显示标签而被检索到，系统后台直接进行同一性处理。当然，为了与传统的规范记录在格式上兼容，也可以保留首选词（preferred name）。

（2）任何属性都可以作为“检索点”，也都可以进行规范控制。书目信息中的所有属性特征，只要有需求，都可以作为“标目”或检索点，提供规范的、基于概念的检索。当然也可以不进行词汇控制，采用一般的全文检索、字符匹配的方式。

（3）规范记录（规范档）不再是含混不清的MARC格式，而是可以用清晰记录语义关系的

RDFS 及其扩展(如 SKOS、OWL)等编码模式(schema)。例如,美国国会图书馆推出 BIBFRAME,专设一种“规范数据”格式,就是以 RDFS 形式表达规范数据,所涉及的概念术语及各种属性关系、约束关系及取值范围等,均能以机器可读的方式代码化,包括各类功能需求模型(如 FRBR/FRAD/FRSAD 等)所规定的各种关系。

(4)万维网的全球一体化环境为分布式规范控制的自动更新和同步机制提供了很好的平台。通过 Web 服务,各类参照关系可以跨域整合和自动服务,实现包括编目和检索过程的各类功能需求。国际组织和各国的国家图书馆通力协作,还可以引入社会化众包模式进行规范档的更新纠错维护,利用大数据分析进行自动的规范术语获取和推荐等。这些都是 MARC 时代根本无法做到的。

只有这种依靠最新计算机网络技术实现自动化的管理,才能将各类编目原则和具体规则(如 RDA/AACR2 等)代码化、语义化,应用并融合到规范控制的整个过程中去,才有可能克服人工流程的各种不一致情况,实现规范控制的最大诉求。

6 万维网规范控制

万维网时代是一个信息严重过剩,而知识十分稀缺的时代,掌控人类所有知识的书目控制理想虽然越来越遥不可及,但也绝非应该被抛弃,反而更加彰显价值,在科研、教育、生产等领域更应得到重视。

对于历史上已经出版的文献,已有 OCLC 这类国际机构以及世界各国的国家图书馆实现了基本的书目控制。当前最大的任务,是将这些书目信息向万维网迁移,使之成为人们随时可用的参考,这就需要为书目数据制订新的、适应互联网时代的格式和交换标准。目前,OCLC 针对 WorldCat 这一全球联合的书目数据库研发出符合 Schema. org 规范的书目数据扩展格式,能够将书目信息以 RDF 形式嵌入到网页中,从而能被搜索引擎所收割和识别。OCLC 还联合美国国会图书馆等机构,研发了规范档的关联数据服务虚拟国际规范档(Virtual International Authority File,VIAF)。另外,美国国会图书馆也开发了符合关联数据规范的 BIBFRAME 书目数据格式,其中规范数据是其四种数据类型之一。此外,还有大英图书馆、德国国家图书馆等一大批国家图书馆都将自己的国家书目发布成了关联数据。这些新的数据规范一方面充分考虑与过去的 MARC 数据兼容,保证书目数据的语义内容能够迁移到新的系统中,另一方面也为未来的书目控制探索了可行的技术方案。

对于万维网时代“原生”的“文献”如何进行“书目控制”,除了国际图联的一些报告,或学者的论文之外,并无系统的研究。万维网时代“文献”的概念已发生了巨大变化,按照“文献是记录有知识的一切载体”[13]的经典定义,它在数字时代可以以任何形式和媒体形态呈现,其负载的内容和载体可以完全分离,甚至“碎片”化、“数据”化了,它还可以同时呈现于用户的任何终端设备上。“书目”的含义也随之发生改变,可以是对任何知识单元的描述,并且不局限于图书馆行业,至少包括博物馆、美术馆、档案馆等在内的所有“记忆机构”都有类似的“书目控制”需求。

这些变化一方面使得“书目控制”的数量有了很大增长,内容类型也大大超出以前的范围;另一方面值得进行书目控制的内容在知识总产出中所占的比重将会越来越小,大量的知识产出由于各种原因不需要或无法进行“控制”,图书馆的书目控制距离涵盖所有知识载体类型的梦想将会越来越远。不严格地说,搜索引擎才是目前互联网信息的最大“控制”者,虽然目前它只具备少量的规范控制功能,但语义万维网技术正在使搜索引擎能够索引知识,谷歌的知识图谱(Knowledge Graph)、Wolfram Alpha[14]等就预示着这个发展方向。

因此,未来的书目控制将只能存在于某些特定的、有规范控制需求的领域,例如科学研

究、工程管理、社会运行、产业经济、教育媒体等,这些领域需要通过付出额外的人力和其他成本来获取一定的有序性,要求越高,成本越大。类似化学文摘社(CAS)这样的基于知识的规范控制,永远是有需要的,只是它主要由人工来完成标引加工的业务模式会发生变化。将来大多数的元数据加工和规范控制工作应该无需专门的编目人员去做,规范信息将越来越多地能够伴随知识的生产、流转等生命周期过程中,由软件或系统自动生成和附加。

语义万维网技术为万维网时代的规范控制提供了原生的解决方案,但如何做却主要不是技术问题,而是一个管理和决策问题,不同的应用领域有不同的需求,资源情况和业务流程也不一样,因此也决定了不同的实施成本,这就带来一个规范控制的"度"的问题,不是说越严格的规范控制就越好,科研成果和文学作品的要求肯定是不一样的,虽然双方都很关注责任者的标注,但在内容揭示方面,对于前者显然希望更准确地揭示(如前面列举的化学文摘的例子),而对于后者,如果我们希望把文学作品里的双关、反讽、隐喻、幽默以及话里有话也标注出来的话,显然失去了欣赏的意义。

从技术的角度,Burners - Lee 提出的关联数据四原则和五星级标准[15]提供了规范控制严格程度的参考;从书目控制角度,规范档的丰富程度也决定了规范控制的"高级"程度。然而,总体上万维网环境下的规范控制只能追求合适,无法追求完美。评价是否合适主要是以能否满足需求为标准,即在多大程度上满足了规范控制在特定领域的功能需求,如前述国际图联功能需求研究报告中总结的查找、辨识、提供情境、证明、选择和探索等,以及更多的本地需求。需要选择哪些属性做规范(即标目),以及是选择控制词表的方式进行严格规范,还是仅仅定义属性元素的定义域和值域,以及数据类型或数据之间的关系,这些都可以由具体应用来决定。这些其实就是 MARC 规范档中所记录的内容,一旦决定,都可以以 RDFS 方式进行形式化编码,使机器可读、可校验甚至可解析参考。

7 一些实例

把书目数据的揭示和服务迁移到互联网上,是近 20 年来图书馆界探索得最多的主题,其中以美国国会图书馆于 2012 年底推出的 BIBFRAME 书目框架格式草案和 OCLC 虚拟国际规范档的尝试最为著名,且影响深远。然而仅多一种兼容过去的书目数据格式是没有意义的。

(1)美国国会图书馆的 BIBFRAME Authority 规范数据格式

书目框架(BIBFRAME)是美国国会图书馆于 2011 年启动的一项研究计划,它的目的是开发一种"适应未来需求"的书目数据格式,即 BIBFRAME,逐步取代 MARC,使书目数据在万维网上被方便地发布和共享。该格式应用了关联数据技术,能够对图书馆及相似机构的各类馆藏资源进行描述和编码,规范数据是 BIBFRAME 四种数据类型之一(其他三个分别是作品、实例和注释)[16]。

书目框架定义的规范数据格式并非要取代其他的规范控制方法,而是作为一种容器,提供一个轻型的抽象层,使规范控制在万维网环境下更加有效地发挥作用。它既要实现传统规范控制对作品、实例及其相关的作者(人物及角色)、机构、主题、事件等要素的规范功能,兼容传统的 MARC 规范档数据,又有许多新的网络化特点,如支持向其他规范数据服务(例如 VIAF 或 DBPedia)的外链、支持规范档的编码描述以及对数据的属性取值提供自动的链接解析校验等。BIBFRAME 的规范控制定义了四个子类:代理(agent)、地点(place)、时间(temporal)和主题(topic),并对它们的编码做出了具体规定(见图 3),MARC 规范记录中的属性描述基本上都能转换成书目框架的 RDF 陈述来表达。

图 3 BIBFRAME 规范控制模型

(2) OCLC 的虚拟国际规范档(VIAF)

VIAF 是美国国会图书馆(LoC)、德国国家图书馆(DNB)和 OCLC 于 1998 年发起的一个规范数据服务研究项目，2007 年法国国家图书馆加入，2012 年成为 OCLC 的一项服务。VIAF 利用了关联数据技术，将各国国家图书馆的规范名称数据集成在一起，提供全球范围的规范数据服务。至 2014 年 7 月，其成员已发展到 29 个国家 34 个机构，包含了来自 30 个国家的 35 个图书馆的数据，还在不断接收新的成员。

虽然分布式计算并不要求数据集中存储，但 VIAF 这种将各成员馆规范数据整合在一起的做法，无疑有利于项目实施和统一管理。互联网环境下非常需要规范数据的统一服务，这对于全球图书馆的数据加工，以及图书馆数据面向整个互联网的开放存取都有巨大好处。OCLC 作为一个图书馆的联合体，其自身并无能力生产数据，但是它可以通过前瞻性的研究开发，将大家的数据整合起来提供更好的服务，WorldCat 是这样，VIAF 也是这样。

作为开放关联数据的一员，VIAF 已能集成其他关联数据(如 DBpedia、Wikidata 等)作为规范数据(见图 4)，而且其所规范的内容也不仅仅是人名、机构名、地名、统一题名、主题等，还有许多其他名称或概念(如虚构人物、动物、国王、主教、圣徒、天使、总统、城市、湖泊、山川等)，它还考虑和采用了标准的名称标识，如 ISNI、ORCID 等，将来的服务也不局限于图书馆领域。截至 2014 年 6 月，VIAF 已有 3 516 万人名数据，543 万机构数据，388 万题名数据和 42 万地名数据[17]。

康奈尔大学发起、多个研究机构参与的 VIVO 项目，看似一个科学家的社会性网络，但实质上是科学家人名的规范控制，它不仅采用了科学家个人、研究机构和专业人员(如图书馆员)合作建立规范数据的模式，而且采用语义万维网技术管理和发布数据，以应用程序接口(API)方式提供规范数据的参考引用服务[18]。另外，近年来有一项由博物馆界发起的“开放规范项目”(Open Authority)[19]，尝试汇集图书馆界与博物馆、美术馆等人类记忆机构的各类资源，利用社会性网络，采用众包的方式，共同开发规范控制服务。维基百科也在实施一个类似的 Wikidata 项目[20]，采用维基百科的方式，将海量的各类事物、概念的名称，以关联数据的方式发布，并支持解析和引证。西班牙格拉纳达大学(Universidad de Granad)基于 Drupal 开发的规范控制 Authoris 系统[21]，支持 MARC 等格式或符合 FRBR/FRAD 模型的数据以关联数据形式发布，并提供较为完善的编辑、引用和发布功能。

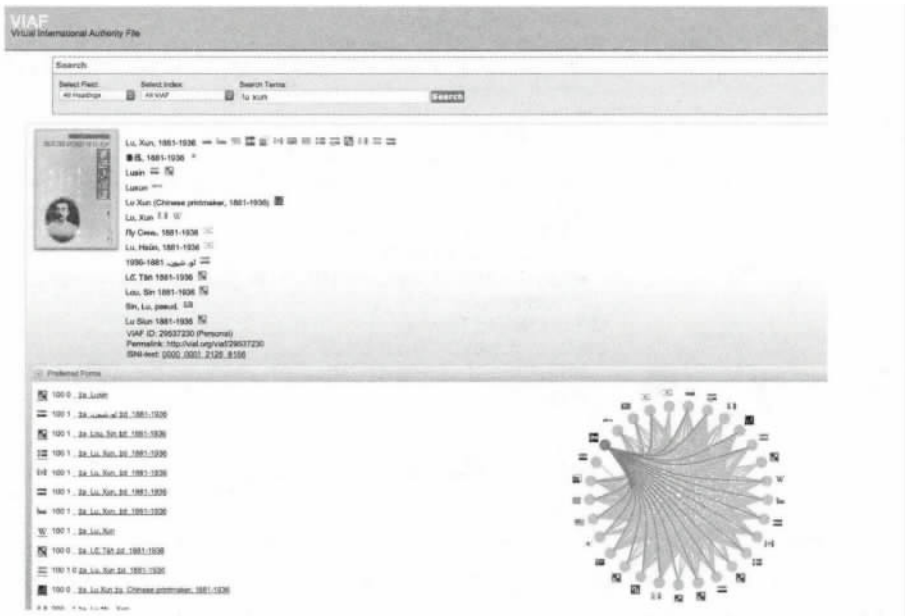

图 4 VIAF 中鲁迅的条目

8 结语

“在今天的环境下，书目控制不能再被看作局限于图书馆目录。”

“书目控制未来将是合作的、去中心化的、国际范围的、基于 WEB 的。”

“单一环境(如图书馆目录)中描述(著录)的一致性，与各种环境间进行连接的能力相比，正变得不那么重要。”[22]

以上论断来自美国国会图书馆 2007 年发布的《书目控制未来报告》，该报告预言了一个新时代的到来，虽然我们刚刚站在门口，但已看到巨大的机会和挑战。上述基于万维网的规范控制努力反映了一种发展趋势，即传统的图书馆知识组织和整序工作，在互联网时代还是有价值的，图书馆行业数百年积累起来的书目控制经验，如果能充分利用好现代信息技术所提供的强大工具，不仅能实现过去没有实现的理想，而且能在更大范围内发扬光大。

参考文献

[1] 高红. 编目思想史[M]. 北京: 北京图书馆出版社，2008. (Gao Hong. A history of cataloging perspectives [M]. Beijing: Beijing Library Press，2008)

[2] 丘东江. 新编图书馆学情报学辞典[M]. 北京: 科学技术文献出版社，2006. (Qiu Dongjiang. A new encyclopedic dictionary for library & information science [M]. Beijing: Science and Technology Literature Publishing House，2006)

[3] Reitz J M. Online dictionary of library and information science [EB/OL]. [2015 - 03 - 31]. http://www.abc-clio.com/ODLIS/odlis_A.aspx.

[4] 黄俊贵. 规范控制概说[J]. 高校图书馆工作，1999(3): 1 - 8. (Huang Jungui. A survey of authority control [J]. Library Work in Colleges and Universities，1999(3): 1 - 8.)

[5] 吴冰，王浩，陆彩云. 现代书目控制理论与实践[M]. 北京：知识产权出版社，2014.（Wu Bing, Wang Hao, Lu Caiyun. The theory and practice of modern bibliographic control [M]. Beijing: Intellectual Property Press, 2014.）

[6] 王松林. 中文编目与 RDA[M]. 海洋出版社，2014.（Wang Songlin. Chinese catalogue and RDA [M]. Ocean Press, 2014.）

[7] 云明向. 论书目控制理论在网络信息资源组织中的利用[J]. 四川图书馆学报，2006(6)：32－34.（Yun Mingxiang. Bibliographical control and network information resources organization [J]. Journal of Library Science in Sichuan, 2006(6): 32－34.）

[8] 顾犇. 关于《书目控制未来报告》草案[J]. 国家图书馆学刊，2008(1)：76－78.（Gu Ben. About the draft of *Report on the Future of Bibliographic Control* [J]. Journal of the National Library of China, 2008(1): 76－78.）

[9] 富平，刘小玲. 中文书目规范控制的理论与实践[M]. 北京：北京图书馆出版社，2007.（Fu Ping, Liu Xiaoling. The theory and practice of Chinese bibliographic control [M]. Beijing: Beijing Library Press, 2007.）

[10] 国际图联书目记录的功能需求研究组. 书目记录的功能需求最终报告[EB/OL]. [2015－01－12]. http://www.ifla.org/VII/s13/frbr/frbr-zh.pdf.（IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records final report [EB/OL]. [2015－01－12]. http://www.ifla.org/VII/s13/frbr/frbr-zh.pdf.）

[11] Tillett B B. Virtual international authority file [EB/OL]. [2015－01－12]. http://www.nl.go.kr/icc/down/060813_3.pdf.

[12] Library of Congress. Bibliographic framework initiative [EB/OL]. [2015－01－20]. http://www.loc.gov/bibframe.

[13] 中华人民共和国国家质量监督检验检疫总局，中国国家标准化管理委员会. GB/T 3792.1—2009 文献著录总则[S]. 北京：中国标准出版社，2010.（Genearl Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. GB/T 3792.1—2009 Bibliographical description [S]. Beijing: Standards Press of China, 2010.）

[14] WolframAlpha [EB/OL]. [2015－01－20]. http://www.wolframalpha.com.

[15] Berners-Lee T. Linked data [EB/OL]. [2015－01－20]. http://www.w3.org/DesignIssues/LinkedData.html.

[16] Library of Congress. BIBFRAME authorities [EB/OL]. [2015－01－22]. http://www.loc.gov/bibframe/docs/bibframe-authorities.html.

[17] OCLC. 2014 annual report to VIAF council [EB/OL]. [2015－01－22]. http://www.oclc.org/content/dam/oclc/viaf/OCLC-2014-VIAF-Annual-Report-to-VIAF-Council.pdf.

[18] VIVO Conference Workshop. VIVO: data integrity and maintenance by example [EB/OL]. [2015－01－25]. http://www.vivoweb.org/files/presentations/12ws5/DataMaintenanceAndIntegrity.pdf.

[19] Phillips L B. Defining open authority in museums [EB/OL]. [2015－01－29]. http://midea.nmc.org/2012/01/defining-open-authority-in-museums.

[20] Vrandecic D, Krotzsch M. Wikidata: a free collaborative knowledgebase [EB/OL]. [2015－02－12]. http://korrekt.org/papers/Wikidata-CACM-2014.pdf.

[21] Leiva-Mederos A, Senso J A, Iominguez-Velasco S, et al. Authoris: a tool for authority control in the semantic web [EB/OL]. [2015－02－13]. http://arxiv.org/ftp/arxiv/papers/1402/1402.2019.pdf.

[22] Library of Congress Working Group. Report on the future of bibliographic control [EB/OL]. [2015－02－25]. http://www.loc.gov/bibliographic-future/news/lcwg-report-draft-11－30－07－final.pdf.

刘 炜 上海图书馆副馆长，研究员。通信地址：上海市徐汇区淮海中路1555号。邮编：200031。
张春景 上海图书馆协调辅导处副研究员。通信地址同上。
夏翠娟 上海图书馆系统网络中心研究开发部高级工程师。通信地址同上。

（收稿日期：2015－03－13）