
图书馆信息资源组织与利用的发展趋势

宋文

1 书目控制的未来

1.1 总体发展趋势

2006年11月，美国国会图书馆组成了一个工作组，研究21世纪书目控制的发展趋势。2008年1月，国会图书馆发布了工作组最终研究报告¹。

工作组预想书目控制的未来将是合作的、去中心化的、国际范围的、基于Web的。这一未来的实现将发生在与私营机构的合作、与图书馆用户的积极合作下。书目数据将从多个来源汇聚，变化将快速出现，书目控制将是动态的，而不是静态的。

报告基于一个关键的前提：当前正处于书目控制和信息存取的关键转折点。图书馆必须以最有效的、合作的方式，最小化书目控制的费用。

报告对书目控制、书目社会、国会图书馆在书目控制中的作用进行了重新界定和定义，从而揭示未来书目控制的趋势和状态。

1.2 重新认识书目控制

书目控制未来工作组在报告中提出应在更广的范围定义书目控制，书目控制应包括图书馆的所有材料，各种类型的用户，各种信息查找场所或情景。

书目控制要处理各种类型的资源：出版发行后被图书馆购买的资源；图书馆为用户联机开通的资源；公开的网络数字资源；图书馆独有的资源等等。

¹ Report of The Library of Congress Working Group on the Future of Bibliographic Control. 2008.

各种书目实践已经增加了资源的类型：图书馆收藏的图书和期刊、档案、期刊论文、博物馆对象和图像等。这些资源逐渐地可以经由 Web 访问，各书目社会分离地管理这些资源不再是所希望的、可持续的或有足够服务能力的。书目控制逐渐地变成跨书目社会的关系管理——著作、名称、概念、对象之间的关系描述。在单一的环境里（如图书馆目录）保持描述的一致性比起在多种环境之间的相互连接已显的不那么重要，从 Amazon 到 WorldCat，到 Google，到 PubMed，到 Wikipedia，图书馆服务是整个相互连接的 Web 服务中的一个节点。在今天网络化的信息环境下，书目控制不再能够被限制在图书馆目录这样一个局限中。

尽管编目在书目控制中仍将继续发挥关键作用，但今天已经有太多的其他数据来源能组织并提供信息的访问。为了利用其他来源的数据，需要将书目控制看作一个分布的活动，而不是中央化的。关于馆藏使用的数据，引文连接，流通和销售数据等，在书目控制领域都是有价值的信息。用户贡献的数据（如评论、排行）能帮助其他用户确定他们所需要的资源。任何电子数据集合，从图书馆目录到全文文献，都能通过自动挖掘的方法挖掘出新的信息。即使那些资源原始不是在图书馆中，但这些系统能看作是为图书馆用户服务的工具。

1.3 重新认识书目社会

书目社会超越了图书馆、出版者和数据库生产者，包括资源的创建者、供应者、发行者、保存者和用户社会，书目社会将跨越机构和国家的界限。图书馆仅仅是从作者到读者的书目信息供应链中的一种连接。它在书目供应链中是独特的，但不是唯一的。所有的参与方都通过书目记录提供有价值的信息。到目前为止，还没有一个强有力的在出版物生命周期里共享书目元数据的传统。未来需要仔细考虑在出版物的各个生命阶段都将增加哪些信息，已有的信息如何能被下面的阶段获取并利用？如何集成各阶段的信息并建立一个集成的基础机制。

在当今快速成长的市场环境，信息访问也能被看作一种商品。许多从前由非营利部门管理的信息资源现在成为了营利经济的重要对象。图书馆需要承认他们是在这个广阔领域中的一个成员，市场条件使得图书馆与商业机构相互作用和影

响。一个例子是在大规模的数字化建设项目中，营利机构正在利用图书馆资源和图书馆元数据。另一个例子是沿着出版供应链中逐渐增加的数据交换，出版者生产的数据成为联机书店的数据基础，图书馆系统连接到那些联机书店是因为传统图书馆书目记录中所没有的信息，如封面图像或用户评论。

书目环境的扩展和演化是基于互联网的，它成为了国际性的。书目控制不再仅限于图书馆、出版者、数据库生产者的领域。各种出版发行的资源，或公开的资源，以及它们的书目数据，与他们当前的和潜在的用户一起能出现在任何地方，而且每个地方都是同步的。

持续的合作共享是未来图书馆成功的一个关键因素。不仅需要图书馆之间的互相合作，也需要与图书馆以外的机构合作。合作能采取多种形式：合并其他来源的数据到图书馆记录中；图书馆能连接到其他机构创建和维持的数据等。这些方法不是相互排斥的，所有可能的方法都将在考虑之列。

1.4 从重新定位国会图书馆看书目控制的未来

国会图书馆从 1902 年开始发行编目卡片为美国国内以及美国以外的图书馆提供服务，这一服务一直延续到现在。

工作组的报告提出，国会图书馆并不是国家图书馆，它提供书目数据服务并没有得到相应的财政支持。在现在图书馆运行环境（包括技术和经济）已经发生根本变化的情况下，国会图书馆继续承担这样的角色是不切实的，也不一定是合适的，尤其是国会图书馆的法定任务是管理好它本身的巨大量的、复杂的内部馆藏、服务和计划。

国会图书馆已经越来越多地利用出版商和发行商的书目数据，这给其他图书馆一个信号，通过利用这些数据，书目记录创建将更有效率，编目员的时间能更多地释放出来，去从事提供受控检索点这样的更高层次的智力工作，帮助用户更好地发现资源。

报告建议国会图书馆、其他图书馆、出版界要分担书目记录创建的责任。国

会图书馆应检查其编目计划和各子计划，对于超出国会馆直接任务的部分，考虑是否可以由其他相关机构分担。报告还建议国会图书馆与 PCC，ARL、专业机构、出版者等合作，去规划这样一个新的责任分担的转变。

1.5 通过书目社会的广泛合作共享，提高书目控制的效率

未来图书馆的书目控制将在广泛合作的基础上，通过增进出版界、书商、图书馆等各种书目社会的书目记录的合作与共享，通过信息资源供应链生产的书目数据的最大程度的利用来提高整体图书馆书目控制的效率。

因此未来图书馆需要能够灵活地接受其他方提供的与图书馆标准不是准确兼容的书目数据；需要分析并修改书目描述或交换标准，保证他们能够支持与出版商和供应商之间进行数据共享；需要与出版者和其他资源提供者以一种良好的工作方式统筹共享数据。例如开发内容和格式标准，使 ONIX 数据自动被 CIP 采用。

1.6 将图书馆力量迁移到更高价值的活动上

图书馆的重点应迁移到向更多的用户揭示图书馆收藏的还未充分展示和利用的稀有的、独特的资料上。

特藏（包括但不限于图书和小册子、档案和手稿、音视频、照片、地图）对学术研究有很大的价值。特藏是一个图书馆独特的标志，往往是社会文化和知识生活的反映。然而对特藏的处理一直赶不上特藏的采购，在各图书馆中有大量的未处理的特藏，这些资料不能通过联机工具发现和获取。即使这些资料在一个图书馆已经被处理了，但这些资料的书目数据以往都是不能共享的。

工作组建议优先保障图书馆稀有的、独特的、和其他特藏资源能够被查找和发现。改进特藏资料的编目，重点放在更高的覆盖率和更大范围的可检索性。开展多馆合作，实现编目记录和规范记录共享，提供对特藏资料的集成访问等等。

1.7 未来书目控制的技术定位

充分认识 WWW 是图书馆技术平台，也是传递图书馆标准的平台，认识到人类不是书目数据的唯一用户，计算机应用系统也以各种方式进行数据交互操作。

1.7.1 Web 将作为书目数据的基础平台，未来书目数据需要一种新的格式

今天，许多信息访问和书目 workflow 已经移动到 Web 平台上，书目数据存储在数据库中，一般仅用于显示和检索，不能与数据库外的服务交互（如连接到全文），不能与 web 资源交互（如地图和参考工具）。

存储在独立的图书馆数据库中的数据经常不暴露给 Web 应用，通用搜索引擎不能检索到，这样使用这些通用 Web 应用工具的用户也不能查找到这些信息。

图书馆社会的数据载体——MARC，是基于 50 年前的老的技术，不能适应当今的编程风格。除了图书馆社会，没有其他社会使用这一格式，严重影响了它作为数据传输工具的作用。图书馆环境以外的书目应用不使用、可能也不兼容 MARC 记录。未来书目数据需要一种新的格式，能容纳和区分专家、机器、用户生成的元数据，包括注释（评论、注解）和使用数据。应允许在不同环境下，有选择的（模块化的）元数据使用，如受控词汇适用于专门领域。由于老的格式的限制，现在的 Z39.2/MARC 不是一种合适未来书目数据载体。

图书馆已经定义了许多标准词汇，如地名，受控词汇和规范目录，帮助图书馆社会创建相互兼容的资源描述。但是有些词汇仅仅作为文本文档使用，经常掩埋在很长的各种标准中，如一般资料标识包含在 AACR 中，MARC 语言代码包含在 MARC 标准中。这些词汇在图书馆社会已经产生了很大的价值，但是不能被其他社会所共享。

文字字符串，如个人和团体名称作为标识符用于显示和数据处理实际上阻碍数据跨语言、跨信息社会的交换。文字字符串作为标识符将款目绑定到了单一语

言上，阻碍了规范文档和使用规范标目的书目文档的国际化。文字字符串可以随时间而变化，数据被不同的应用使用的越多，采用独立于语言的标识方案就越重要。更理想些，这些方案应也能提供情景信息，也能连接到额外的关于数据元数据和它的含义的信息。

1.7.1.1 建立一个更灵活、可扩展的元数据载体

认识到 Z39.2/MARC 不再适用将来的目标，LC 和其他有兴趣的社会一起建立并实施书目信息载体，能够表达图书馆的全面数据，并能实现图书馆社会间，相关社会间的数据交换。

1.7.1.2 将图书馆标准集成到 Web 环境中

用机器可读的、机器可操作的格式表示图书馆标准，尤其是那些应用于 Web 平台的标准。

通过标准的注册或 Web 站点发布，使标准能被所有 Web 应用使用。

LC 应开始将其管理的词汇表迁移到 Web 服务友好的平台，允许文档被下载合并进入其他应用。这些词汇包括书目记录中的语言和地理区域代码、资源格式代码等。

1.7.1.3 扩大标准标识符的使用

LC 应生产其维护的标准的 Web 为基础的标识符。

1.7.1.4 期望的结果

图书馆书目数据将从封闭的数据库转变到开放的 Web 平台，书目记录是可编址的，其格式能容易地集成进 Web 服务和计算机应用。这样能够使图书馆更好地利用网络数据资源，利用 Web 上各种数据源之间的联系的优势。

与广泛的兴趣方协调，尤其是图书馆以外的书目数据创建者，图书馆社会将开发一个记录载体，能够无缝地与图书馆数据和图书馆系统集成，能同时被图书馆和其他社会使用。载体格式将支持各种书目控制时间和资源类型。

由图书建立的词汇将用于 Web 发现，容易地被应用开发者重用。词汇将用注册和其他结构管理，使得更新更快比用中央化的管理的目录。知识组织系统将促进多语言词汇的转换和交叉。

网络环境下的所有数据点将被明确地标识，主要使用统一资源标识符（URI）。数据点的注册将包括含义和使用信息。图书馆社会将共享作者、著作和其他书目数据受控元素的标识符，能够使数据在不同的使用社会交换，能允许数据元素按照有关社会的特殊需要以变化的方式显示和索引。

1.7.2 标准

实施有效的发现、请求、解析和传递，需要系统之间容易地通讯。

图书馆系统逐渐多地进行国际性的数据交换，与非图书馆合作伙伴，如 Google, Microsoft 和 Amazon 进行数据交换。数据沿着出版/书商/图书馆/集成商链条被重用。数据沿着发现到发现的链条更多地被使用，促进更新型的服务。吸收合并多个图书馆系统的数据的新的发现环境正在出现。传统图书馆标准（Z39.x/MARC/AACR2）不能提供与其他信息社会进行数据交互的最好方法。

今天的元数据环境包括 AACR2/RDA, MARC21, MARC XML, MODS (Metadata Object Description Schema), Dublin Core, ONIX (Online Information Exchange format) 等，检索协议包括 Z39.50, SRW/U, MXG (Metasearch XML Gateway)，需要与开放检索和其他协议一起工作。标准正在指数级数地增长，正式注册的和事实上的标准。标准呈几何级数增长分散了国家标准化机构的精力，对从业者是一个混乱，对开发者是一个苦恼。图书馆社会重点需要确认和寻找切实可行的解决方案，防止无效的主张和专业化意识成为发展的主流。

1.8 未来的书目社会定位

促进其他用户提供的信息合并进入图书馆资源描述中，揭示和利用资源间的各种关系，实现 FRBR 框架的潜能。

1.8.1 为今天和明天的用户设计

连接合适的外部信息与图书馆目录；开发系统能够关联到评价数据，如评论和等级；鼓励图书馆系统连接到合适的用户追加数据（如 Amazon.com, LibraryThing, Wikipedia）。

集成用户贡献数据到图书馆目录：开发图书馆系统能接受用户和其他非图书馆的数据。

进行基于数据计算的应用研究。利用馆藏和流通信息指引用户到达他最有用或最有潜在兴趣的文献；鼓励进行计算技术研究，支持书目控制，包括书目数据创建和书目数据的服务。

1.8.2 期望的成果

图书馆书目数据将被使用在一个更加广阔的环境中，图书馆与非图书馆应用之间的互操作将增长和改善。

图书馆目录将作为相互关联的发现工具阵列中的一个有价值的组成部分。图书馆通过用户贡献数据增强资源的发现和评估能力。

1.8.3 实现 FRBR

自从 1998 发布 IFLA 书目记录功能需求的最终研究报告后，FRBR 框架一直作为重新概念化书目数据和书目关系的催化剂。

最近图书馆和其他领域的建模实践（FRBR, CIDOC Conceptual Reference Model, indecs Metadata Framework）提供了成熟的模型，使这一领域再次成为关注的重点。同时，Web 环境中面向资源的体系结构已经使图书馆意识到提供简单的 Web 协议和方案访问数据资源的益处。上述两个方面预示 LC 未来的方向是 Web 为基础的，网络层次的书目控制。工作组预想书目基础结构，其中关于实体的数据（如著作、地点、任务、概念、年代）以议定的方式编码，通过议定的 Web 协议由应用和其他服务使用。LC 和图书馆社会需要寻找到在 Web 上发布其

长期投入的语义数据的方法。

一些系统试验用 FRBR 的 work 概念在用户端聚类资源，这些试验证实了 FRBR 模型 work 定义层的价值，但是仅试验了 FRBR 模型的一小部分。FRBR 模型对编目实践和机读记录的作用还没有得到广泛的研究。目前还没有交换 work 为基础的数据的方法，也没有编目规则支持 FRBR 模型。

维持现状的结果：

图书馆社会正在将它的未来的编目规则放置在一个框架上，但这一框架只是刚刚开始探究。在模型还未被全面测试形成书目数据前，FRBR 只能看作一个理论模型，它的实际应用和附带化费还是未知的。

1.8.3.1 建议

建立 FRBR 测试计划，LC,OCLC,IFLA 工作组，有代表性的系统供应商，识别那些协议是支持 FRBR 所必须的；建立和议定 schema 用于 work 为基础的数据交换；以上述活动为基础，公布并评估 FRBR 的应用。

1.8.3.2 预期成果

FRBR 的研究、细化和验证将提供一个更强有力的框架用于资源描述和访问规则的建立，在未来支持更宽范围的相关检索。最终的产品将是一个具有明确定义的元素及其关系的书目环境，用于各种书目控制情形。

1.9 其它新类型（互动、可视化、“多媒体”、虚拟现实）

Pages 2.0

可使用户对图书馆所保存的珍贵的图书在电脑上进行虚拟现实的翻阅，用户可以放大细节，阅读和聆听专家的点评，同时还可以保存

和分享用户自己的笔记。

Pages 2.0 Windows XP 用户可用用宽带免费下载 .NET Framework version 3, 在 Windows Vista 或 Windows XP SP2 下运行

2 规范控制的未来

工作组收到了大量关于规范控制的现状和未来发展的信息。所有信息一致地表示图书馆和图书馆用户对于编目记录采用国会图书馆提供的当前的、有效的、无二义性的检索点已感到很厌烦。

有很多关于机器检索能力的改进、数据库能力的增长消除了对名称、主题概念的规范需要的抄作，但公众的证词和可用的证据显示事实并不如此。虽然关键词检索提供用户一种非常有用的检索能力，但它们不是受控词表的令人满意的替代品。的确，许多机器检索技术依靠现有的规范标目，即使这些标目不被明确地显示出来。

规范标目对用户发现和确认资源是至关重要的，但规范标目的建立极大地增加了编目的费用。尽管可以通过共同承担规范记录建立在一定程度上降低费用，但规范记录创建的需要受资源本身和知识资源增长率所决定。任何人可能是一个作者，新作者增长很快。为继续进行有效的规范控制，需要采取一系列策略。一个策略是采取自动的方式协助进行规范控制，例如协助区分相似的作者。还可以采取让出版者和作者参与到作者的标识过程中来的办法。

主题分析，包括内容分析、主题标目和分类号的分配，是编目的核心功能。主题标目是按论题聚类的基础。LCSH 的复杂性造成了主题标目创建和使用的困难。

书目数据的国际化需要不同语言标目相同，虚拟国际规范文档是朝向这一方

向的一个脚步。最后，需要创建一种数据结构，使用中立的、不依赖语言的标识符标识术语和标目。

维持现转的结果：

规范控制被限制在图书馆应用环境中，经常是建设良好或大型的系统有能力应用规范数据。

规范控制的侵蚀（Erosion）将会阻碍用户快速地检索到所需的信息。关键词检索变的越来越流行，非文本关键词、英语以外的语言的著作正变的更少被访问到，甚至是根本无法访问。

工作组建议增进规范数据创建的合作。LC 和 PCC 应确定一种方法促进各方广泛的参与规范数据的创建、完善、维护；开发工具促进规范记录创建，更好的在图书馆 workflow 中集成、共享规范数据；LC、PCC、OCLC 研究一种方法激励对国军爱书目数据库和规范数据库中新的规范记录的建立和规范记录的更新维护。

工作组建议提高对已分配的规范标目的重用。LC、图书馆社会、图书馆系统供应商、出版者应调研名称规范的衔接，识别各种情景下的需求，如图书馆、出版、仓储管理等。LC 应联合其他社会一起工作解决作者和其他创建者的识别问题、映射问题，调研合作的可能性；将 LC 名称规范文档作为 Web 开放资源，提供下载或通过 Web 服务接口连接。

实现规范文档的国际化。LC、OCLC 和国家图书馆应进行更积极有效的的国际共享的规范文档的建设；采用统一的方法连接国家和国际规范文档。

预期目标：

图书馆、图书馆系统、图书馆以外机构的规范数据的共享将增长，任何处理编目数据的人都可使用规范数据。通过将同一实体规范次数的最小化节约规范的花费。同一实体名称从一个系统到另一系统的交换将变的更为简单。访问规范数据将没有任何障碍，使用数据的障碍也将最小化。

通过广泛的相关方的合作与协调，新的合作关系将建立，这将促使实体的规范数据创建工作流的效率和最小化重复工作。

国际共享的规范文档将增进对非英语文献的访问，包括对非罗马字符和文字文献的访问，将促进国际信息和数据的共享。

德意志图书馆、美国国会图书馆和 OCLC 联机计算机图书馆中心联合开发了一个个人名称的虚拟国际规范文档(VIAF)，它连接世界各国国家书目机构的规范记录，并将免费在网上提供检索。

美国国会图书馆和德意志图书馆的规范文档和书目文档被用来创建初始的 VIAF，包括 600 多万个名称和 50 多万个连接。

VIAF 项目的长远目标是要将许多国家图书馆和其他重要来源的规范名称结合起来，形成一个共享的全球规范服务。

目前对万维网之未来的一些设想考虑使用本体来使得万维网更为智能化，可以用机器来自动处理。如果与其他来自文摘索引服务、档案馆、博物馆、出版社等来源的可控词汇和规范文档结合，VIAF 可以成为“语义网”的一个基本组成部分[2]。图书馆现在面临一个很好的机遇，可以在这个方面作出较大的贡献，并促使这个前景成为现实。对于这个达成共识的前景来说十分重要的是，VIAF 应该是在网上免费提供服务的。

3 主题控制的未来

3.1.1 为使用和重用优化 LCSH

主题分析是编目的核心，LCSH 在提供受控主题访问著作方面提供了巨大的价值。LCSH 被广泛地采用，常常是图书馆目录中的唯一主题搜索点。但从管理和自动化角度看，LCSH 是非常繁琐的。LCSH 中许多已知的缺陷是任何主题词表固有的，主题词表要包括知识创建的整个范围，而不是离散的主题领域。新增主题术语是基于文献。过去，如果 LC 不拥有某一主题的材料，该主题将不会进

入词表。随着 SACO(the subject authority component of the PCC)的实施, 这一问题已经得到一些改变, 其他图书馆也能提交推荐的主题术语。

其他问题是由于 LCSH 已经经过了很长时间, 词表最初不是按照叙词表设计的, 叙词表的关系(等价、相关和等级)现在作为标题被吸收进词表。术语有时是过时的, 对无经验的用户也是不直观的。LC 更新主题术语, 但是 LC 以外的更新结果经常是任意的和预料不到的。

预先协调过的主题字符串, 结合了学科主题、地理主题、年代和类型等, 是一个耗时和复杂的过程。由编目员仔细加工的主题字符串经常不能被用户和参考馆员理解。

LC 主题规范记录是 MARC 格式, 限制了它在图书馆以外的使用。

3.1.1.1 维持现状的结果

LCSH 的复杂性、更新的任意性和复杂的使用条件, 已经对编目员和目录用户产生了一个负面的结果, 减少了其在图书馆以外的使用。

非主题性和非等级层次的组织, 使 LCSH 系统化、协调化地进行更新变得困难。

构造主题词的复杂性导致主题词易出错, 也由此造成受控词汇的不一致性, 影响相关资料的检索, 最终是受控词表的目的未达到。

3.1.1.2 建议

转换 LCSH 成为一个工具, 提供更灵活的方法创建和修改主题规范数据: 使得 LCSH 开放可用的, 被图书馆和非图书馆各方; 提供 LCSH 字母顺序的排列, 能够使其定制成专题叙词表; 增加明确的关系和参考在 lcsch 术语、lcc 和 ddc 号之间;

LC 应与寻找合作伙伴开发 LCSH,LCC,DDC 这些受控词表的能力, 描述或确定产品或方案, 使得这些术语能在一个更可访问的环境下提供给更广泛的用

户；评估 LCSH 支持分面浏览发现的能力。

鼓励其他受控词表的应用，鼓励与其他受控词表的相互参照。

LC 和其他主体词表的提供者：提供与其他受控主体词表，如 Mesh，国家农业叙词表，Sears List of Subject Headings,³⁷ and the Getty Art & Architecture Thesaurus 的参照，使得词汇可交叉检索和可互操作；