

图书馆目录平台化的技术实现方案研究

夏翠娟（上海图书馆）

摘要 图书馆目录的平台化对于图书馆在互联网时代的转型有着至关重要的意义。欧洲数字图书馆、美国公共数字图书馆等国外大中型图书馆，主要通过框架化的数据模型、基于关联数据技术的数据开放关联、开放应用程序接口、泛在的嵌入式应用、用户协同创新的模式这五种技术方案实现图书馆目录的平台化，各有侧重，可供国内图书馆参考借鉴。

关键词 平台化 开放数据 关联数据 应用程序接口

DOI: 10.13663/j.cnki.lj.2015.09.004

From Portals to Platforms: The Implementation Technologies for Libraries

Xia Cuijuan (Shanghai Library)

Abstract The platform of library catalog is very important to the transformation of a library in the Internet age. Europeana, the Digital Public Library of America and other large libraries have been implemented the platform, mainly by designing the framework of data model, publishing linked open data, developing open application program interface, deploying embedded mobile apps everywhere, and creating user collaborative communities. These approaches can be learned and adopted by the domestic libraries.

Key words Platform, Open data, Linked data, API

0 前言

近年来，随着“互联网思维”的影响和信息产业界的推动，图书馆目录正在寻求从门户到平台的转型。《图书馆论坛》2015年第7期发表的《从门户到平台——图书馆目录的转型》一文中这样描述门户和平台的差别：“门户的典型特征是‘推’，推送资源、推送服务，用户只是被动地接受，信息的流动是单向的，用户的作用被弱化。而平台的典型特征是‘拉’，不仅贡献资源和服务，还能建立网络和社区，把有相同目标的用户‘拉’到一起来，为他们提供创造新价值的平台；同时与其他的网络和社区建立广泛的联系，充分利用外部资源和服务；对自有资源、外部资源、资源提供商和用户进行深度整合，形成完整的产业链条，吸纳用户贡献知识，让信息无限流通、在流通中产生新价值。”^[1]“门户是用来访问的，而平台是用来构建的”。^[2]门户是静态被动的，平台则可为共

享和重用、创新和协作提供无限的可能。

图书馆目录是图书馆各种书目数据的总称，是图书馆最核心最宝贵的财富。图书馆目录的平台化（以下简称“平台化”）是图书馆在互联网世界占据一席之地、发挥应有的知识传播和信息交流作用的需要。国外一些图书馆较早地认识到平台化对于图书馆在新的信息环境下生存和发展的重要性，已经在平台化上取得了一定的成果和经验，国内也有一些图书馆或多或少地在谋求和实施面向平台化的转型。本文提出了平台化的五项技术要求，通过调查图书馆界的典型案例，探索针对这五项技术要求的实现方案，以期为国内图书馆实现平台化提供参考。

1 平台化的技术要求

1.1 框架化的数据模型

平台化要求系统要能支持跨领域跨系统的

资源深度整合。在国外已经形成一种共识:图书馆、博物馆、档案馆甚至美术馆都被看做同一类的文化继承机构,有着相似的资源管理和需求,它们之间的界限越来越模糊。无论是哪一类文化继承机构,如果要与其他类似的机构进行基于数据层而非系统层的资源整合,需要设计新的或利用已有的框架化的数据模型,来满足不同来源、不同种类的数据混搭需求。

所谓框架化的数据模型,是指不单为某一资源、某一机构而设计的数据模型,而是考虑到多种资源、多个机构的总体需求,设计一个抽象的框架,允许不同机构针对不同种类的资源,在此框架的基础上进行扩展或限定,填充应用规则细节,以达到适应具体应用需求的目的。而基于同一模型的应用,能够方便地实现基于数据语义层的互操作。容纳 2000 余个文化继承机构数据的欧洲数字图书馆(Europeana)所设计的欧洲数据模型(Europeana Data Model, EDM)即是一个较为成熟的、应用广泛的、框架化的数据模型,美国国会图书馆(LOC)为了取代图书馆传统的机读目录格式,与语义网技术研发公司 Zepheira 合作开发的、尚未最终完成的书目框架(BIBFRAME)也是这样一个框架化的数据模型。

1.2 数据的开放关联

平台化要求数据与更多的外部数据源相关联,同时利用外部数据源的数据丰富自有馆藏,能方便地被其他机构和用户共享和重用,这就要求数据要以标准化的方式开放获取,采取规范的数据格式输出数据,也能便捷地整合外部数据源的数据。关联数据技术作为一种从数据底层即可实现语义开放关联的实现方案,被众多图书馆应用。

关联数据建立在现有的成熟的网络技术框架之上,利用 HTTP URI 标识和定位数据对象,利用 RDF 抽象模型来描述数据对象的详细信息,利用标准化的 RDF 序列化格式对数据进行编码,利用 SPARQL 协议和查询语言来进行数据读写,可实现跨系统跨领域的数据交互和数据整合。各国国家图书馆包括英、美、德、法、瑞典、芬兰、西班牙、韩国等,还有最大的图书馆目录数据提供商 OCLC 以及欧洲数字

图书馆,都利用关联数据技术实现书目数据的开放关联。

1.3 开放的应用程序接口(API)

平台化要求能基于现有的资源和服务,构建新的应用,创造新的价值。应用程序接口(Application Programming Interface, API)就是这样一种可被广泛采用的技术途径。API 允许资源和服务提供者将自己的数据封装为可被程序调用的接口,让计算机程序之间方便地交互,让资源和服务像搭积木一样自由组合,从而构建新的应用。欧洲数字图书馆和美国公共数字图书馆(The Digital Public Library of America, DPLA)就提供了强大的 API 来供应用程序访问其后台千万级的资源。上海图书馆也以多样化的 API 构建了上海市公共图书馆的技术支撑与服务平台。

1.4 泛在的嵌入式应用

在互联网时代,为用户提供随时随地的服务是基本的要求。随着社会化网络的风靡和移动设备的普及,图书馆实施了“到用户所在的地方去”的战略,不再局限于在门户网站上被动地等待用户来访问资源和服务,而是将图书馆的服务嵌入各种 SNS 平台、网络社区、移动 App 和 Web App,主动为用户提供无处不在的服务。除了开发各种用于移动终端的、用户喜闻乐见的 Apps 外,还在社交媒体如 FaceBook、Twitter、新浪微博上开通账号,一些有实力的图书馆还可以利用这些社交媒体所提供的开发平台,开发自己的服务,例如国内图书馆的微信订阅号、公众号、微店等。

1.5 用户协同创新的模式

平台化要求能将志同道合者拉到一起来,形成人和思想的网络。10 年来,Web2.0 技术的兴起在全世界范围内引发了 Lib2.0 的风靡,用户创建内容(UGC)的模式也深刻地影响了图书馆的信息交流和生成方式。图书馆本身可以从这样的模式中受益,许多图书馆开始吸引用户参与资源管理和服务的工作流程,比如根据用户反馈改善服务的质量,用户驱动的采购(PDA),多用户协作标引平台等。同时这样的模式也能使用户获益,它为用户提供突破个人知识局限和思想藩篱的途径,聚集的不仅是跨领

域的资源，还有群体的智慧。例如全球最大的博物馆、图书馆、档案馆、美术馆联合体史密森（Smithsonian）的转录平台，就是一个吸引用户参与贡献，多用户协同工作的成功范例。

2 平台化的案例研究

2.1 设计框架化的数据模型实现资源的深度整合

欧洲数字图书馆和美国公共数字图书馆之所以能够容纳多来源、多种类、多格式的大规模数据资源，和它们所采用的数据模型 EDM 是分不开的。EDM 被设计成为一个跨领域的语义互操作框架，兼容多种领域标准，如博物馆的 LIDO、档案馆的 EAD、图书馆的 MARC。同时这些不同的领域对 EDM 的发展也做出了贡献。作为一个优秀的框架化的数据模型，EDM 具有以下特点：

开放性：其词表复用了来自 DC、FOAF、GEO、FRBR、ORE、RDFS、SKOS 中的术语，与 GeoNames、DBPedia 等数据集相关联。

重用性：EDM 最初设计时就考虑到要尽可能地便于重用，其设计宗旨是：所有限制它重用的约束都是需要摒弃的。已在美国公共数字图书馆中作为核心数据模型得到重用。

可扩展性：有两种方式可以对 EDM 进行基于特定领域需求的扩展，一种是对 EDM 中一些已有的类和属性定义更具体的约束和规则，使其具象化；另一种是增加新的类和属性，使其继承 EDM 中已有的类和属性，这样面向特定领域应用的扩展方案被称为应用纲要（Application Profile）。

众包：在 2015 年至 2020 年的未来开发计划中，特别强调对“众包”的支持，将基于“Web 注释数据模型（Web Annotation Data Model）”扩展，提高用户参与度，如丰富已有的元数据，添加基于上下文语境的语义标注信息。^[3]EDM 是一个不断发展的数据模型，其维护和升级的过程也尽可能地公开透明，让更多的合作者、业界专家、最终用户都参与进来。

2.2 应用关联数据技术实现数据的开放关联

大英图书馆、美国国会图书馆、欧洲数字图书馆、OCLC 是图书馆界应用关联数据技术的先锋和典范。

大英图书馆于 2012 年将书目数据发布为关联数据，采用了将 MARC 数据转换为 RDF 格式发布的方式。在本体词表设计上，选择了大量非图书馆界的本体词表来建立大英图书馆术语（British Library Terms, BLT），如 W3C 的 Organization 本体和 Event 本体、FOAF、GeoNames 地理空间本体等。利用这些本体词表，实现了书目数据中不同实体间的内部关联，以及与 Geonames、DDC、LCSH、VIAF 等外部数据集的开放关联。^[4]

美国国会图书馆主要是将各种规范数据发布为关联数据，包括 LCSH、LC 名称规范档、LC 分类法等，是作为数据提供者对其他数据集提供服务。目前，关联到国会图书馆规范数据的数据集众多，根据 OCLC 的关联数据应用调查，美国国会图书馆的关联数据服务站点 id.loc.gov 的日访问量达 10 万次^[5]。美国国会图书馆正在进行的 BIBFRAME 项目试图设计一种基于关联数据的数据模型来取代 MARC，并计划于 2016 年将 MARC 数据转换为 BIBFRAME 格式发布，使书目数据从封闭走向开放。

欧洲数字图书馆的关联数据先导项目始于 2012 年初，目前已有 2000 万种以上的多媒体资料的元数据发布为关联数据，以批量下载、内容协商、SPARQL 端点、API 等方式提供开放获取。采用 EDM 为关联数据模型，在众多数据提供者的数据之间建立了关联，也与欧洲办公室发布的数据、还有 GeoNames 中的地名、DBPedia 中的人物、GEMET 叙词表中的概念、古代时间词表中的时间等外部数据建立了关联。^[6]

OCLC 从 2012 年开始将 WorldCat 的书目数据发布为关联数据，采用由 Google、Yahoo!、Yandex 等三个搜索引擎巨头发起并设计的 Schema.org 作为本体模型和词汇来源。通过在 HTML 网页中嵌入 RDF 三元组的方式来实现关联书目数据的发布，目前 WorldCat 上的条目网页嵌入了 RDFa（RDF attribute）和 Microdata 这两种数据格式。因而其书目数据能够在网络环境中被搜索引擎发现。^[7]

2.3 利用开放 API 实现资源的共享、重用和整合

欧洲数字图书馆、美国公共数字图书馆、澳大利亚图书馆、新西兰数字图书馆均提供了

API^[6]。通过调研这四家图书馆的 API 发现,要使 API 达到良好的可用性,需做好以下几个方面:

公开访问:API 需在 Web 上公开访问,一般是以 Restful Web Services 的方式,允许 API 调用程序在 HTTP URL 中传入既定的参数,获取返回的数据,返回数据一般为 XML、JSON、JSON-LD 等标准的、可被大多数程序语言处理的格式。

权限控制:为了更好地掌控 API 的使用,一般要求开发者注册账号,获取 API Key 来控制 API 的访问。

开发文档:提供完善的开发者指南文档,对 API 访问方式,参数,返回数据格式,可获取的数据内容范围及内容结构做出详细的说明。

代码示例:提供调用 API 的代码示例,或代码库,帮助开发者快速地掌握 API 的使用方法。欧洲数字图书馆甚至提供一个 API 控制器(console),允许开发者在未注册之前对 API 的参数和返回结果进行测试和试用。

用户社区:以 BBS、邮件列表等方式为开发者提供交流和反馈的平台。

2.4 开发泛在的嵌入式应用实现随时随地的服务

美国公共数字图书馆、欧洲数字图书馆、澳大利亚图书馆、上海图书馆等国内的众多图书馆都开发了大量的嵌入式应用。

澳大利亚国家图书馆开发了一个 Wordpress 插件,帮助用户方便地在自己的博客文章中嵌入图书馆提供的报纸文章,并链回图书馆网站。TroveNewsBot 是澳大利亚图书馆在 Twitter 上的一个自动咨询账号,随机发布图书馆数据库中的报纸文章并回答读者的咨询。这个账号和美国公共数字图书馆的账号 DPLABot 互动,找到双方发布的有关联的内容,建立链接,还能自动地从历史数据库中找到与最近的新闻相关的文章并发布,激活历史馆藏,同时将读者引回图书馆的网站。

上海图书馆除了开发移动终端的各种 Apps 外,还在微信平台建立公众号,提供查询、续借等服务,将图书查询服务嵌入微信和支付宝的城市服务,与水电煤缴费、看电影等大众服务并列。

2.5 建立用户协作平台实现增值和创新

史密森(Smithsonian)的转录平台^a支持

上传或从对象库中导入扫描的原始图片资料,包括手写信件、笔记、日记、钱币、动植物标本等。系统针对不同的资源对象,提供相应的元数据著录界面,对元数据元素的取值进行一定的控制。请用户协助著录,对已有的元数据进行补充,对机器无法识别的手稿类资源进行全文抄录。系统实时显示每一个项目的进度,包括有多少人参与,已完成多少页,还有多少页未完成,吸引大量有相同爱好和专长的志愿者参与,相互协作,完成著录标引任务。以留言的方式,支持用户之间相互的交流和反馈。在质量的控制上,采取了一套简单有效的“抄写——复审——核准”流程,任何注册用户都可以参与复审,使得不同用户在相互协作的同时,还能彼此促进,获得知识的提升。用户贡献的内容也得到很好地利用,可转成 PDF 下载并加盖用户贡献勋章,同时生成元数据记录并建立索引,可被其他用户检索。该平台上线一年内,上传图片 12 万页,参与人数达到 5000 以上,来自 172 个国家,在标引速度和质量上取得了较好的效果。^[9]

3 结语

在平台化的实施上,欧洲数字图书馆是较为优秀的案例,从数据模型的设计到关联数据应用、开放 API 的构建、嵌入式应用的布局和对用户参与业务流程的支持,都进行了全面的规划和实施。而美国公共数字图书馆将不同来源的数据深度整合,基于多功能的 API,向互联网贡献高质量的数据,成为数据基础设施的一部分。澳大利亚国家图书馆和新西兰数字图书馆也在平台化上迈出了关键性的一步,在开放 API 和嵌入式服务上表现优秀。史密森在对多用户协作的支持上做出了大胆的创新,提供一个平台,把文化继承机构、领域专家和用户凝聚在一起,实现信息的增值和知识的增长。国内图书馆目录要实现从门户到平台的转型,可借鉴欧洲数字图书馆、美国公共数字图书馆等典型案例,参考本文所提出的 5 项技术方案,逐步实施。

a <https://transcription.si.edu/>

(下转第112页)

员。研究方向：图书馆管理。 E-mail: jhshen@lib.tongji.edu.cn 上海 200092

杨 锋 女，硕士，同济大学图书馆，馆员。 上海 200092

张更平 女，硕士，同济大学图书馆，助理馆员。 上海 200092

(来稿日期：2015-04-07)

(上接第22页)

参考文献

- [1] 夏翠娟, 吴建中. 从门户到平台——图书馆目录的转型[J]. 图书馆论坛, 2015(7):1-7.
- [2] Tim Sherratt.From portals to platforms: building new frameworks for user engagement.[2015-07-12]. <https://www.nla.gov.au/our-publications/staff-papers/from-portal-to-platform>.
- [3] Valentine Charles, Antoine Isaac. Enhancing the EuropeanaData Model.[2015-07-18].<http://pro.europeana.eu/publication/enhancing-the-europeana-data-model-edm>.
- [4] CorineDeliot.Publishing the British National Bibliography as Linked Open Data[EB/OL]. [2015-07-16].http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf.
- [5] Jeff Young, etc.OCLC Linked Data Research[EB/OL].[2015-07-17]. <http://www.oclc.org/research/themes/data-science/linkedata.html>.
- [6] <http://labs.europeana.eu/api/linked-open-data/data-structure/>.
- [7] TedFons, Jeff Penka, Richard Wallis. OCLC ' s Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web[EB/OL].[2015-07-20]. http://www.niso.org/apps/group_public/download.php/9408/IP_Fons-et-al_OCLC_isqv24no2-3.pdf.
- [8] DigitalNZ.Use the DigitalNZ APIs to access files from across the NZ cultural, education, and government sectors [EB/OL].<http://digitalnz.org.nz/developers>.
- [9] Kristin Conlin. Increasing Access: The Smithsonian Transcription Center[EB/OL].[2015-07-10].<http://siarchives.si.edu/blog/smithsonian-transcription-center-and-william-healy-dall>.

夏翠娟 女，上海图书馆（上海科学技术情报研究所）系统网络中心研究开发部，高级工程师。研究方向：数字图书馆。E-mail: cjxia@libnet.sh.cn 上海 200031

(收稿日期：2015-08-17)

(上接第32页)

- 图书馆学刊, 2014(3): 47-52.
- [9] 3375. 01 State library board. [2015-01-17]. <http://codes.ohio.gov/orc/3375.01>.
 - [10] 李松武. 认真制定执行事业单位章程推进事业单位法人治理结构建设[J]. 中国机构改革与管理, 2013(7): 97-99.
 - [11] 肖容梅. 我国公共图书馆法人治理结构建设现状与分析[J]. 国家图书馆学刊, 2014(3): 22-28.
 - [12] 李国新. 我国公共文化机构的法人治理结构试点[J]. 图书馆建设, 2015(2): 4-7.
 - [13] 罗珊珊. 纽约公共图书馆的法人治理结构[J]. 图书与情报, 2014(2): 17-19.
 - [14] 李松武. 建立健全法人治理结构创新事业单位体制机制[J]. 中国机构改革与管理, 2012(1): 63-66 .
 - [15] California Library Services Board Members [EB/OL]. [2015-01-17]. <http://www.library.ca.gov/LOC/board/index.html>.
 - [16] 金武刚. 大英图书馆的法人治理结构[J]. 国家图书馆学刊, 2014(3): 41-46.
 - [17] 张岩. 我国公共图书馆理事会制度建设面临的几个难点问题[J]. 国家图书馆学刊, 2014(3): 29-33.
 - [18] 无锡新区图书馆[EB/OL]. [2015-01-17]. http://www.wxqlib.com/news_detail.cfm?uuid=E0FF8764-776E-4B98-98FE5CF83EC4C7AF&list_num=20¤t_page=1&msgtype=8.
 - [19] 市图新馆落成启用[EB/OL]. [2015-01-17]. http://tourism.kcg.gov.tw/tw/index.asp?au_id=45&sub_id=46&id=2072.

顾晓光 北京大学信息管理系，博士研究生，馆员。 E-mail: gxg@pku.edu.cn 北京 100871

刘兹恒 北京大学信息管理系，教授，博士生导师。 北京 100871

(收稿日期：2015-04-10)