

图书馆集成融汇服务研究*

李春旺

(中国科学院国家科学图书馆 北京 100190)

[摘要] 在介绍集成融汇服务发展背景基础上, 文章重点讨论图书馆集成融汇服务的目标、类型以及实现方法, 并分析了集成融汇技术在图书馆应用现状、特点及发展趋势。作者认为, 开放数据、Web 服务、表示层组件构成图书馆集成融汇主要对象, 开放接口、结构化数据以及支撑工具是实现集成融汇服务的前提条件和必要保障。

[关键字] 数字图书馆 集成融汇 技术 应用现状

[分类号] G250.76

Study on Library Mashups

Li Chunwang

(National Science Library of Chinese Academy of Sciences, Beijing 100190, China)

[Abstract] Based on introducing to the background, this article discusses goals, types and requirements of library mashups. Mashups can be divided into three main types: data, functionality and presentation. Open APIs, structured data and support tools are requirements for mashup implementation. Finally, the author analyzes present situations, features and trends of mashup application in the field of the library.

[Keywords] Digital library Mashup Technology Current situation

1 集成融汇服务背景

近年来, 集成融汇服务 (Mashup) 在图书馆领域得到迅速发展, 相关的背景原因是多方面的。首先, 建立在单独馆藏资源之上的信息服务很难满足用户日趋复杂化的需求, 对第三方组织、机构、个人资源的集成利用成为当前用户需求发展的一个新趋势。其次, 基于图书馆端系统的服务越来越不应用户需求, 纽约公共图书馆 (NYPL) 用户调查发现, 大部分年轻学生平时主要使用的是社会网络空间 (如 Facebook、mySpace)、个人电脑桌面系统 (如 Mac OS Dashboard、Google Desktop) 等^[1], 如何将图书馆的资源与服务嵌入到用户所在环境中, 成为数字图书馆建设面临的新问题。第三, Web2.0 时代, 创新 (Innovation) 与竞争 (Competition) 成为影响图书馆发展的两大要素^[2], 用户不再是单一的网络信息消费者 (Read-only), 他们同时还是信息生产者、网络服务构建者 (Read-write-and-participate)^[3], 进而成为图书馆员的合作对象、竞争对象。因此, 支持用户自服务、支持用户参与网络构建等成为图书馆服务建设的一个新目标。集成融汇技术为应对以上挑战提供了新的解决方案, 因此它受到图书馆界的广泛关注。

所谓集成融汇, 就是融合多个不同来源的数据资源与服务创建出新的内容复合的服务产品过程。图书馆集成融汇服务的目标就是要充分利用已有的资源与服务, 通过集成创造新的服务, 并支持这种服务向用户工作流的嵌入。关于集成融汇概念、组织机制等, 在文献^[4]中有专门论述。本文将重点讨论图书馆集成融汇对象类型、实现方式以及图书馆融汇服务发展特点等, 希望为相关研究与

* 本文系中国科学院 2009 年新增能力项目“知识服务集成融汇关键技术研究”成果之一。

建设提供借鉴。

2 集成融汇对象类型

根据融汇操作目标对象不同，集成融汇可分为：数据融汇、功能融汇、表示对象融汇^[5,6]。

2.1 数据融汇（Data Mashup）

数据融汇是从多个开放数据源获取相关数据、动态构建出新的内容综合的数据对象，以满足用户复杂的信息需求，这些数据可能来自本地馆藏，也可能来自第三方组织结构或个人。数据融汇又分两种：第一种是简单混合（Combination），将多个不同来源、具有不同属性的数据按主题、时间、位置等进行组织、排列。iSpecies.org就是这样一种生物物种信息搜索引擎，它针对用户提交的物种查询请求，从Wikipedia获得该物种的名词术语信息，从NCBI获得相关基因组信息，从GBIF（Global biodiversity information facility）获得该物种全球分布信息，并从Yahoo!获得物种图片信息、从GoogleScholar获得相关学术文献信息等，然后对这些信息进行简单合并（Join），构建出一个内容综合的结果对象。第二种是分析汇聚（Aggregation），不但动态获取多来源数据，而且借助一定工具、算法从这些数据对象中抽取相关信息片段、发现隐藏在背后的知识，创建新的信息对象。例如：全球疾病警报系统HealthMap.org，它从全球多家权威医疗卫生网站采集数据，包括世界卫生组织、EuroSurveillance、Google News、ProMED、Moreover Technologies、Wildlife Disease Node等；然后，利用自动翻译技术实现英语、西班牙语、法语、葡萄牙语、俄语、汉语等多语种资源的统一融汇，抽取疾病信息；最后，按疾病种类、发病国家地区、发病时间等属性进行统计分析，形成分析结果并与Google地图融合完成可视化呈现。

2.2 功能融汇（Functionality Mashup）

功能融汇是指从不同来源选择执行特定功能操作、具有标准接口的多个Web服务，将其按一定工作流程进行组合构建出新服务的过程。功能融汇的主要特点是：使新系统建设工作站在巨人肩膀上，避免一切从头开发，缩短新开发周期，保证图书馆能对用户迅速变化的需求作出敏捷反应。例如：天气预报网站Weather Bonk（<http://www.weatherbonk.com/>），它通过调用HostIP.info获取当前用户所在地区信息，调用Google Maps查询用户相邻城市信息，调用Weather.com获得用户所在城市及相邻城市天气预报信息，调用WeatherBug.com获得来自附近地区天气监控摄像头的实时图像信息，最后，通过调用Google Maps API将天气预报信息、实时监控图片信息等标注到地图上，支持用户利用地图浏览、查询天气信息。

可视化融汇（Visualization）是功能融汇的一种主要形式，即把信息检索服务、分析计算服务等与可视化交互服务融合，创新图书馆信息揭示、利用方法。目前常用的可视化融汇方法有多种，其中：①时间融汇法（Timeline、Timeplot、Timegrid）是将信息对象按时间先后标注在同一个时间轴上，指导用户按时间浏览相关信息。②空间融汇法（2D Map、3D Map）是将信息对象按其所在的地理位置标注在2D或3D地图上，便于用户利用地图功能对信息进行查询、定位、浏览。③展览馆法（Exhibit）则模拟现实生活中展览馆信息组织、揭示方法，在不借助复杂的数据库视图技术以及服务器端处理技术前提下，将数据对象按Timeline、标题列表、地图可视化等方式进行组织，支持用户对

综合数据资源的分类检索、浏览。④分类图表法（Chart）是将信息对象与饼图、柱图等图表可视化工具集成。⑤主题空间法将信息按主题聚类为空间区域，以相同主题信息数量确定空间大小，根据主题相关性确定主题区域布局，例如Newsmap（<http://newsmap.jp/>）。

2.3 表示对象融汇（Presentation Mashup）

表示对象融汇是集成融汇应用的最简单形式，它通过选择、组合多个不同来源的表示层对象构建出一个新应用，使用户在一个页面里就可以利用多个网站、应用系统的资源与服务。表示对象即表示层组件，通常表现为一段HTML代码片断，主要由两部分组成：一是定义组件在浏览器中的布局风格、外观显示属性等；二是定义组件内容输入输出属性，通常又分为数据型与操作型两种。数据型组件需要指定一个或多个数据源，它们可以是RSS feed、可返回结构化数据的REST查询请求、或者是二进制数据文件URL等。同时，需要定义数据显示方式，对应输出不同组件类型，如数据列表（Data Grid Widgets）、分类图表（Chart Widgets）、数据地图（Maps Widgets）、时间轴组件（Timeline Widget）、展览馆组件（Exhibit Widget）等。操作型组件通常表现为连接某一操作功能的交互界面，需要定义操作方法、Web服务调用接口、应答呈现方式等，操作方法通常采取内嵌脚本语言方式实现。目前最常用的操作型组件是检索框（SearchBox），例如，欧洲图书馆TEL（The European Library）发布Mini Search Box^[7]，支持第三方用户在自己的页面中嵌入TEL检索框。

表示层组件是一段HTML代码片段，无需编译，就可以被拷贝、粘贴到第三方HTML网页中，充分体现了代码重用的思想^[8]。同时，表示层组件也可以定义成一个XML对象，通过在第三方融汇应用服务器上注册，实现向该应用环境的嵌入。目前，很多基于Ajax技术的用户桌面系统都支持表示对象的开放注册与个性化定制，如Google、My Yahoo、Netvibes、Afrous等。

3 集成融汇实现

NISO认为^[9]，一个好的数字资源对象只有具有完整的上下文关系和开放接口，才能支持向用户个人工作流的嵌入，从而达到被广泛利用的目标，而要支持用户参与的融汇组织则需要相关工具支持。

3.1 开放接口是实现服务融汇的前提

目前，Web服务开放接口标准主要有两类：第一类是面向专业开发人员的重量级接口SOAP/WSDL，可以支持复杂的信息交换及基于工作流的服务组合（Service Composition），但对开发人员的计算机程序设计技能要求比较高。第二类是轻量级接口REST/WADL^[10]，采用HTTP方法（PUT、GET、POST、DELETE）实现对资源的CURD（Create、Read、Update、Delete）操作，常用REST接口有SRU、OAI、OpenURL、RSS等。对于集成融汇应用而言，REST比SOAP拥有更好的发展前景^[11]。首先，SOAP采用基于远程过程调用式交互模型，要求调用者熟悉服务描述与发现机制（WSDL、UDDI），使其只适合用户已知、共用一个数据模型的闭合应用环境；其次，SOAP接口过于复杂，对接口方法调用顺序、参数类型等有较强约束，难以适应开放的、分布式的、可伸缩的Web需求。REST不但接口简单，而且具有可伸缩、可信赖等优点。目前，越来越多的服务供应商都提供REST接口，如Amazon、Google、Yahoo等，SOAP/WS-*支持者也开始分享REST经验。

AJAX是一种支持客户端浏览器服务融汇的接口规范。它利用JavaScript将Web服务嵌入到网页

中，采用HTTP协议由浏览器动态连接指定Web服务，并对异步返回的结果信息进行处理、显示。数据描述规范通常采用JSON（JavaScript Object Notation）^[12]，JSON采用独立于语言的文本格式编码，既支持结构化又易于读写；采用压缩方式传输信息，占用带宽小；支持多种语言（包括Actionscript、ColdFusion、Java、Javascript、Perl、PHP、Python、Ruby等）对数据结构的直接解析与映射，代码可重用性强，减少了XML解析方面的性能问题和兼容问题。因此，在客户端融汇应用中，JSON比XML更有优势。

3.2 结构化组织是实现数据融汇的基础

数据质量是应用集成融汇发展的一个重要因素。为实现对半结构、无结构数据的集成融汇，需要对数据进行结构化转换，通常有两种做法：一是在融汇过程中借助屏幕抓取、信息抽取等技术，将非结构化数据转换成结构化数据，很多数据融汇工具都包括屏幕抓取模块，如 OpenKapow、SnapLogic。二是在融汇实现前对相关数据源进行结构化处理，这也是开放数据运动（Open Data）的一个努力目标。W3C Linking Data^[13]是一个有代表性的项目，它借助信息抽取、RDF、URI 等技术对 DBpedia、FOAF 等开放网络资源进行结构化描述，建立不同数据源之间的语义关联，支持第三方用户对这些数据的开放访问与语义集成。另一个项目是 Connexions (<http://cnx.org/>)，它按知识单元（Module）组织、描述、存储教学信息，每个 Module 是一个最小主题信息对象，内容采用基于 XML 的结构化描述，封装标准接口，支持 Module 之间的关联。采取像维基百科那样的共创共享机制，信息消费者可根据自己教学需要从 Connexions 选择相关主题 Module，将其按一定流程进行编排、组合，从而构建出个性化教学课件。

3.3 工具支持是用户自我融汇的保障

从 Web 服务接口规范复杂程度看，基于 SOAP 和 REST 的融汇实现都需要有计算机程序设计技能做支撑，而基于 AJAX 接口规范的表示层融汇虽然相对简单，但让科研用户自己手工创建一个新的融汇服务仍然不是件容易的事情。因此，设计开发相关工具对于推动用户参与的融汇服务建设非常重要。关于集成融汇工具，请参见本专题相关文献。

4 集成融汇技术在图书馆的应用

4.1 通过融汇地图服务创建图书馆资源揭示新方法

首先，将图书馆信息与地图服务融汇，支持用户利用地图定位图书馆并获取相关服务信息。Lewis & Clark图书馆系统(LCLS)包括研究图书馆、公共图书馆、学校图书馆、专业图书馆等几百家成员馆，为了方便读者利用这些图书馆，LCLS创建一个基于Google地图的图书馆定位服务^[14]。在Google地图上标出各个图书馆所在位置，当用户点击某一图标时，系统自动弹出关于该图书馆的通信地址、网站地址、联系电话、开放时间等基本信息，支持用户利用地图功能搜索图书馆。HCPLC^[15]、Minnesota州立大学图书馆^[16]等也都建立了类似服务。

其次，将图书馆的信息资源与地图服务进行融汇，构建基于地图的资源揭示、检索、浏览服务。McMaster大学图书馆为了更好地揭示馆藏航空摄影照片资源，他们按地理位置将照片标识在Google地图

上, 并对拍摄年代进行区分, 支持用户通过地图查询、浏览图片信息, 成为倍受用户欢迎的图片发现工具^[17]。Auckland大学Stuart Lewis通过将ROAR和OpenDOAR采集到的数据与Google地图融汇, 创建一个全球开放存取资源导航网站(<http://maps.repository66.org>), 以地图方式揭示1000余个数据库、1600余万条开放数据在全球的分布情况, 并提供基于仓储软件的资源统计服务。

4.2 利用融汇技术扩展 OPAC 服务

包括: 将 OPAC 检索服务封装成标准融汇组件 Widget, 支持向用户环境的嵌入; 将 OPAC 书目记录信息与地图 (Map)、时间轴(Timeline)、展览馆(Exhibit)等可视化呈现方法融合, 创建书目信息揭示新方法^[18]; 为 OPAC 动态集成第三方相关资源与服务, 如封面图片、书评信息、目录或章节信息、定价及销售信息等, 提升书目信息揭示深度; 为 OPAC 集成社会化服务功能, 使 OPAC 成为一个社会网络平台。例如: John Blyberg 利用 OPAC 数据库提炼出 4 个服务^[19]: 图书馆最热门图书、最新图书、当前所借图书以及预约图书, 并用 Google Gadgets API 将这四个服务封装成四个融汇组件, 通过在 Google Fusion 服务器上的注册, 用户就可以在个人 iGoogle 页面中定制以上服务。另外, 希腊 Panteion 大学采用 Web2.0 技术对 OPAC 系统进行改造, 集成融汇社会标签、用户评价与分级 (Rank)、基于标签相似性搜索等新功能^[20]。而美国印第安纳州立大学 Variations/ FRBR^[21]项目以 FRBR 数据模型改造传统 OPAC 组织体系, 支持音乐数字图书馆系统融汇多类型相关资源, 构建复杂的 OPAC 信息对象。

4.3 融汇技术促成图书馆资源管理与服务理念的转变

首先, 集成融汇技术的应用促进图书馆服务从主要依赖本地馆藏资源转向对第三方开放资源的充分利用。安大略剑桥公共图书馆有一个非常漂亮的新书展示台服务 (Book Cover Carousel)^[22], 其新书信息来自本馆 OPAC 数据库, 图书封面图片信息则取自 Syndetic Solutions^[23]。Syndetic 是一个信息服务公司, 已经建立了 400 万种图书封面图片数据库, 每个封面包括大、中、小三种不同尺寸的图片。剑桥公共图书馆虽然没有单独建立自己的封面图片数据库, 但通过动态调用 Syndetic 的资源, 使自己的新书展示服务提供高清晰图片信息。其次, 集成融汇技术促进了图书馆服务的重用。Purdue 大学为其学位论文数据库封装 SRU 接口, 提供 feed 数据服务及 XHTML 页面服务。之后, 在构建教师个人简历页面时, 动态调用该接口服务获得每个教师曾指导过的研究生及学位论文信息, 并动态将其嵌入到教师简历页面中, 避免人工填写与维护, 当新的学位论文进入数据库后, 指导教师简历页面的相关信息也会自动更新^[24]。

4.4 通过创建表示层融汇组件, 支持图书馆服务向用户所在环境的嵌入

将图书馆的资源与服务封装成用户可以开放访问的表示层组件, 并提供这些服务向用户环境嵌入的解决方案, 支持科研用户在社会网络空间、用户桌面系统等个人环境中调用图书馆的服务。例如, 纽约公共图书馆 (NYPL) Homework NYC Widgets 项目^[25]对图书馆基本服务进行封装, 创建了一系列融汇组件, 包括: 支持检索、评价、推荐等功能的书目服务 Widget, 集成馆藏书目资源、Google、Weikpedia 等数据库资源的跨库检索 Widget, 定题检索服务 Widget, 个人网络书签 Widget, 教师咨询服务 Widget 等, 并聘任专门的社区服务馆员指导用户在个人信息环境中调

用以上服务。与纽约公共图书馆相比, Texas 大学图书馆的融汇服务类型更加多样化^[26], 它不仅包括可嵌入到 iGoogle、Netvibes 等工具中的表示层组件, 还包括嵌入到用户端浏览器中的工具条、嵌入到办公软件中的插件、嵌入到 iPhone 等移动设备中的组件等。JISC 资助 ICO3 开发了机构仓储服务组件, 包括: 上载文献、查询已发布文献的授权、查询基金提供者对项目的授权政策、查询 OPENDOAR、查询 ROAR(The Registry of Open Access Repositories)等, 支持科研用户从个人常用信息环境中访问机构仓储系统的服务^[27]。国内, 国家科学图书馆于近期发布了融汇服务目录^[28]。

4.5 融汇多方相关资源, 建立面向特定研究领域、特定主题的综合信息服务

基于集成融汇技术, 建立面向特定主题的综合信息资源与服务体系成为当前数字图书馆建设的重要特点。例如Georgia州立大学人权数字图书馆CRDL(<http://www.civilrightslibrary.org>), 它不但集成众多图书馆、档案馆、博物馆收藏的有关人权方面的文献资源, 而且集成广播、电影、电视公司收藏的相关影音资料, 同时还集成了故事、评论、人物介绍、多媒体等网络开放获取信息, 并利用集成融汇技术对多类型资源进行融合组织, 建立基于时间轴TimeLine、GoogleMap等多种信息揭示方式, 为人权教学与研究提供内容综合、上下文关联的服务。旧金山现代艺术博物馆也建立了一个综合融汇服务系统Explore Modern Art (<http://www.sfmoma.org/explore>), 它在馆藏资源基础上集成了多媒体藏品讲解资源、在线文档记录信息、公众活动事件信息等, 采用可视化方法建立交互展现空间, 支持基于上下文的信息发现与展示。

5 结语

除了数据资源融汇外, 图书馆集成融汇服务还包括人员融汇、组织机制融汇, 即建立区域内、领域内相关组织机构之间的合作机制, 支持跨组织边界的不同角色人员组成服务团队, 针对用户需求合作建立数据融汇、服务融汇生态环境, 整体提升信息资源保障与服务能力。正如Rick Luce所言^[29], 未来图书馆之间需要建立“云合作”(Cloud Collaborating)机制, 只有这样, 才能够全面推动资源云、服务云建设, 保证图书馆的服务可以被方便嵌入到用户环境并成为用户科研工作流程中不可或缺的一部分, 并在支持科技创新中发挥重要作用。

参考文献

- 1 New York Public Library. A decentralized approach to homework help by public libraries. [2009-06-28]. http://www.ims.gov/applicants/samples/AdvDigitalRes%20Narrative%20Example_NLGOLS.pdf
- 2 Michael S. Institutional innovation, mashup & library future. The Mashed-Up Library (OCLC Symposium), June 28, 2008. [2009-06-28]. <http://www.oclc.org/us/en/news/events/presentations/>
- 3 Harder G. Connecting the dots: social software and the social nature of libraries. Feliciter. 2006;52(2):54-5.
- 4 李春旺.集成融汇概念、组织机制与应用.现代图书情报技术,2008(12):22-26.
- 5 Dion H. Is IBM making enterprise mashups respectable? June 18th, 2006. [2009-06-28]. <http://blogs.zdnet.com/Hinchcliffe/?p=49>
- 6 Andy D. Mashup basics: three for the money. Network Computing, September 5, 2007. [2009-06-28]. <http://www.networkcomputing.com/showitem.jhtml?articleID=201804223>

-
- 7 Mini search box - The European Library. [2009-06-28].
http://www.theeuropeanlibrary.org/portal/organisation/services/minisearchbox_en.html
 - 8 Web widget - Wikipedia, the free encyclopedia. [2009-06-28]. http://en.wikipedia.org/wiki/Web_widget
 - 9 NISO Framework Working Group. A framework of guidance for building good digital collections. [2009-06-28]. <http://www.niso.org/publications/rp/framework3.pdf>
 - 10 Marc J. H. Web application description language (WADL). November 9, 2006. [2009-06-28].
<https://wabl.dev.java.net/wabl20061109.pdf>
 - 11 Paul P. Second generation Web services, 2002,2. [2009-06-28].
<http://webservices.xml.com/pub/a/ws/2002/02/06/rest.html>
 - 12 RFC4627: the application/json media type for JavaScript Object Notation (JSON). 2006,7. [2009-06-28].
<http://tools.ietf.org/html/rfc4627>
 - 13 W3C SWEO. Linking open data. [2009-06-28].
<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
 - 14 LCLS member libraries - Google maps. [2009-06-28].
<http://maps.google.com/maps/ms?ie=UTF8&hl=en&msa=0&msid=109072570655115484151.000463ab9b38bc023ee65&ll=38.872325,-89.946442&spn=0.277984,0.617981&t=h&z=11>
 - 15 Hillsborough county public library cooperative: library locations.
<http://www.hcplc.org/hcplc/liblocales/>
 - 16 The mashed-up library (OCLC Symposium). June 28, 2008. [2009-06-28].
<http://www.oclc.org/us/en/news/events/presentations/>
 - 17 McMaster Libraries. [2009-06-28]. <http://library.mcmaster.ca/maps/airphotos/1997.htm>
 - 18 Michael V. Using Google maps as an interface for the library catalogue. Library Hi Tech,2008,26(1):33-40
 - 19 Go-go Google Gadget!. [2009-06-28]. <http://www.blyberg.net/2006/08/18/go-go-google-gadget/>
 - 20 Dimitris G, Constantia K, Christos P. Enhancing library services with Web 2.0 functionalities. [2009-06-28]. <http://www.dcu.gr/dcu/Documents/documents/enhancing-library-services-with-web-2-0-functionalities/en/attachment>
 - 21 Variations/FRBR: variations as a test bed for the FRBR conceptual model. [2009-06-28].
<http://www.dlib.indiana.edu/projects/vfrbr/>
 - 22 Hot titles carousel. [2009-06-28].<http://www.cambridgelibraries.ca/hot/carousel.cfm>
 - 23 Bibliographic content and data for libraries and booksellers - Syndetic solutions. [2009-06-28].
<http://www.bowker.com/syndetics/>
 - 24 Michael W. Using SRU to create dissertation mashups. ETD 2009, Pittsburgh, PA, June, 2009
 - 25 New York Public Library. A decentralized approach to homework help by public libraries. [2009-06-28].http://www.imls.gov/applicants/samples/AdvDigitalRes%20Narrative%20Example_NLGOLS.pdf
 - 26 Library widgets - University of Texas Libraries. [2009-06-28].<http://www.lib.utexas.edu/tools/>
 - 27 JISC Repository Aggregator Project. [2009-06-28]. <http://www.rwidgets.co.uk/wiki/doku.php>
 - 28 国家科学图书馆融汇服务目录. [2009-06-28]. <http://crossdomain.las.ac.cn/mashup/mashup.html>
 - 29 Rick L. Making a quantum leap to eResearch support: a new world of opportunities and challenges for research libraries. 2008,12. [2009-06-28].<http://www.arl.org/bm~doc/ff08luce.pdf>

(作者 E-mail: licw@mail.las.ac.cn)