

数字学术与公众科学:数字图书馆新生态

——第十三届数字图书馆前沿问题研讨班会议综述和思考

□谢蓉* 刘炜

摘要 图书馆数字技术应用向纵深发展,正呈现三个新的发展方向:一是从数字化到数据化;二是云端化和大数据之后带来的智慧化;三是全方位的网络众包和大众参与。在上海刚闭幕的第十三届数字图书馆前沿问题研讨班(ADLS 2016)上,首次提出“数字学术和公众科学”的主题,连同数据化和智慧化,这三方面的趋势正带来数字图书馆的新变化和新生态。

关键词 数字图书馆 数字学术 公众科学 数字人文

分类号 G250.76

DOI 10.16603/j.issn1002-1027.2017.01.001

图书馆数字技术应用向纵深发展,已呈现三个新的发展方向:①从数字化到数据化,馆藏对象深入到数据层面正在成为图书馆新的业务趋势;②云端化和大数据之后必然带来的智慧化;③普遍 2.0 之后造成全方位的网络众包和大众参与。这是在上海刚结束的第十三届数字图书馆前沿问题高级研讨班(ADLS 2016)传达出的讯息。

本届论坛以“数字学术与公众科学:数字图书馆新生态”为主题,重点探讨了“数字学术与数字人文”“公民科学与大众参与”“关联数据与语义技术”和“开放数据和开放知识”等四个分主题,这些主题为首倡于 1994 年的“数字图书馆”概念带来了新的内容和新的热点,给当今处于“复合型图书馆”阶段的图书馆实践带来新的动力,并指明了发展方向。在短短两天的研讨时间里,会议容纳了 26 个发言,晚餐期间还举办了头脑风暴和专家问答,密集的概念轰炸和观点碰撞让与会者大呼过瘾,原计划 150 人的研讨班吸引了 240 余位关注前沿的同行与会。

1 从数字化到数据化:数字图书馆向知识服务迈进

“仅仅数字化是不够的,数字图书馆应该将工作的中心转向‘数据化’,数字图书馆的下一步是建设数据图书馆。”来自加州伯克利大学东亚图书馆的林

海青先生如是说^[1]。本届论坛开篇的第一个主旨报告——张晓林研究员的“超越数字图书馆:基于数据与分析的知识服务”就对这一观点直接做出了呼应。

所谓数据化,是指以数据为基本单元组织信息并提供服务的方式,区别于传统的以文献(包括网页、电子书这类数字文献)为基本单元的方式,这里的“数据”,是指在万维网上能够独立存在,或经过一定格式的数据编码,使计算机程序能够处理或“读懂”其含义的数据,并非我们通常所说的只能通过人来理解和处理的数据。数据化的主要特点是资源对象的细粒度和语义化,以及其格式是对于内容的描述而不只是形式或结构上的描述,顺应了万维网从 Web of Document 向 Web of Data 的发展趋势。本次会议清华同方公司介绍了他们在碎片化和精准化知识协同服务方面的最新进展,也恰好印证了这一趋势^[2]。

图书馆的数字资源管理问题还没有得到很好解决,仿佛一夜之间,大家都开始谈论数据服务,业界有人诟病这又是一轮炒概念、赶时髦现象。然而我们认为这是一个必然趋势,因为这一趋势并非来自于图书情报行业本身,而是来自我们所处的环境和我们所服务的对象。整个科学研究都在向“数据驱

* 通讯作者:谢蓉,ORCID:0000-0002-2243-2125, xr@suibe.edu.cn。

动的研究”发展,正呈现“所有的科学都在变成数据科学”这一无法回避的现象。随着近年来数字科研(eScience)和数字人文(Digital Humanities)的兴起,大量的科学研究都在变身为数据密集型研究,产生形形色色的“数字×××学”或“计算×××学”,例如“数字文学”“计算语言学”“计算历史学”“计算生物学”等等,提供基本研究方法和工具的“数据科学”应运而生而且得到快速发展。这一现象被美国计算机科学家、图灵奖的获得者吉姆·格雷(Jim Gray)总结为科学研究的“第四范式”^[3]。

图书馆开展数据服务起因于欧美各国近年来对政府基金项目申报要求提交数据管理计划的规定,许多大学和研究机构都要求图书馆员在科研人员申报项目时提供必要的帮助,以符合编制数据管理计划的要求。部分研究型图书馆走得更远,直接参与了其所属机构数据管理系统的研发、管理和运营。复旦大学张计龙副馆长专门讨论了在构建科研大数据管理平台方面的做法和思考,并特别介绍了美国科研数据行业对“数据科学家”的迫切需求和众多iSchool纷纷开设专业培训,对数据化趋势起到了推波助澜的作用。上海外国语大学任树怀馆长也介绍了以大学发展战略为指导构建数字学术服务平台的心得^[4]。

数字人文是目前图书馆为数字学术开展服务的一个重点领域。早期的谷歌等公司转换的数字图书资源被并入Hathitrust和美国国家公共数字图书馆(DPLA)等新的数字图书馆服务体系中,其它图书馆和博物馆、美术馆、档案馆等人类文化记忆机构也纷纷成立数字人文研究中心,由于它们拥有海量特色馆藏,经“文本化”和“数据化”之后,目前已成为各相关领域数字人文研究不可或缺的基础设施。数字人文研究和服务在我国刚刚兴起,本次会议分别有武汉大学王晓光的“面向数字人文的文化遗产图像深度语义组织研究”,北京大学朱本军的“数字人文:图书馆实践的新方向”、广西民族大学欧阳剑的“面向数字人文的数据融合研究”、上海图书馆夏翠娟的“以数据服务为基础的数字人文平台建设方案研究”等报告。他们分别结合各自的项目,生动地介绍了一批成果,另有来自北京水晶石公司的VR技术总监魏记先生专门就数据模型化和可视化做了报告,都展示了目前国内数字人文研究和应用的较高水平。

然而从行业层面来看,图书馆开展数据服务并未成为其理所当然的业务。要确立在数据管理方面的地位,体现图书馆在传统文献、数字资源之外,对于数据管理也能做出独一无二的贡献,还需要更多的图书馆做出更大的努力。

2 从自动化到智慧化:实现随时随地无所不在的嵌入式服务

智慧图书馆的含义通常有两个:一是指采用无线射频(RFID)或其它自动感知等物联网技术之后实现图书馆服务的自助化(无人化)和智能化,通常是指实体图书馆的智慧化;二是通过“大数据分析”,使大数据变成“聪明数据(Smart Data)”,发挥大数据的价值,使图书馆提供更加个性化的服务。数据分析既可以应用于优化传统服务,也可以用于优化数字服务,因此可以认为这是复合型图书馆的智慧化。而本次会议还涉及另一种“智慧”:通过对资源内容进行语义标注,使其“自带智慧”,变成“聪明内容(Smart Content)”,这才是真正意义上的数字图书馆智慧化。

来自上海交通大学的孙翌介绍了一个典型的实体图书馆智慧型发展之路(他称为“智能图书馆”)^[5],综合采用了多种技术,包括自动感知和数据分析处理等,是第一类智慧图书馆的升级版。阿法迪公司的雷向欣博士也提供了一个结合用户行为数据分析的智慧图书馆解决方案,打破了人们对该公司只作RFID硬件技术供应商的固有认识,具有很强的可操作性^[6]。重庆大学杨新涯馆长展示了一个新上线的图书馆门户系统^[7],不仅有许多2.0功能,而且融合了大量的数据分析,能够提供多种聚类、推荐,这可算作第二类智慧图书馆系统。来自美国肯特州立大学的曾蕾教授分享了一个智慧数据应用于人文研究的案例,可以为图书馆行业所借鉴。

关联数据作为语义万维网的一种简化实现,是目前最常见的和最简单的“聪明内容”,已开始科学数据管理、STM(科学、技术、医学)出版、政府信息公开和企业数据管理等领域得到了普遍关注和初步应用,被认为是互联网Web 3.0的发展方向。本次会议有多篇关联数据技术和应用方面的报告,可以说是国内关联数据领域最新成果的一次大检阅。林海青提出关联数据是对网络上的信息进行规范控制的最好方法^[8],郑巧英利用元数据和本体技术对

多种类型资源的资源深度整合提出了方案,胡小菁详细介绍了图书馆第一个应用关联数据作为书目数据格式的最新进展,夏翠娟则提供了一个研究型图书馆开发数字人文关联数据应用的完整案例,陈涛博士专门针对政府开放数据提出了关联数据的解决方案,张永娟对涉及“聪明内容”的语义出版现状进行了概述。

会议最有意义的内容之一应该算是曾蕾和范炜所做的关于“关联数据能力指标(CI4LD)”的介绍^[9]。这是一个国际专家小组最新的工作成果,把学习关联数据有关的知识分解为具有层次的各种指标,参照“成就标准网络(Achievement Standards Network, ASN)”中的《图书馆能力指标》^①,将所有相关知识点进行分类发布,这样就可以指导人们进行对照学习,或评估自己在关联数据方面已经具备的知识和能力。发布 CI4LD 的网站目前已移交由都柏林核心元数据组织(DCMI)进行维护^[10],能力指标文档的中文翻译也已在官网上发布^[11]。本届会议的一个重要成果是将由上海图书馆建立和维护一个中文版 CI4LD 网站,与英文版网站一样,支持能力指标的进一步修订扩充,以及将更多的中文关联数据资源关联上去,供人们学习和参考。

“聪明内容”不仅是有语义的内容,而且是“有情境(Context)”和“可行动(Actionable)”的内容。上海图书馆刘炜副馆长在闭幕报告中展望了机器智能时代图书馆数据服务的挑战^[12],认为当前语义技术的应用还只是“智慧图书馆”万里长征的第一步,将来经由深度学习而来的机器自动标引和知识组织能力将完全能够取代人类,知识组织的大量劳动密集型工作可以交给机器来完成,但是如果我们现在不能给机器提供足够深度学习的算法和模型,即提供情境描述的结构和行动模式的话,机器是不可能进行深度学习的。这也是目前利用语义技术模拟真实世界的最重要的意义所在。

3 从专业化到公众化:以社群力量推动数字学术

Web 2.0 之后,网络不仅是传播信息的渠道,而且是人际交互的平台;不仅是探索知识的宝库,而且发展成连接万物、实现知识创造的虚拟空间。

据估计,美国人每天花在社交网络脸书的时间大约有 40 亿分钟,相当于 7500 人年。他们不论做什么,有意义还是无意义,他们的行为模式本身,就

具有巨大的商业价值,都是对脸书的贡献。目前我国已有 7.1 亿网民,其中手机网民 6.56 亿,他们每周上网时间大约为 26.5 小时^[13],如果能把他们的时间组织和利用起来,或者在满足他们上网目的的同时顺便实现一些其他目标,则将产生巨大的财富。实际上已经有了大量这样的案例,美国社交网络专家克莱·基舍(Clay Shirky)将这一现象称为“认知盈余”^[14]。

将人们的“认知盈余”应用于科研领域,就形成了所谓“公众科学”。公众科学(Citizen Science),又称公民科学、大众科学、群智科学(Crowd Science)、社区科学(Community Science)、公众参与式科学研究(Public Participation in Scientific Research),它是指公众作为研究者参与到科学活动中的过程^[15]。高深的科研工作往往被认为是一种高度的专业化工作,必须经过长期严格的训练才能进行,而现在转变为人人能参与,并不是说科研事业本身降低了难度,而只是其方式过程发生了改变,降低了参与门槛而已。基于数据的科研过程在基于网络的管理平台上,有可能剥离许多数据采集、处理、分析、归纳整理的步骤,把人类擅长的“劳动密集型”工作通过分发众包,从而实现大众参与并推进科研创造。南京理工大学的赵宇翔教授对公民科学现象有着深刻的理解,他总结公众科学通常有如下特点:①群体性;②协作或竞争性;③任务众包;④开放性等^[16]。公众广泛的参与也是数字学术的一个特征。

当一部分“数字公民”徜徉于虚拟世界不亦乐乎的时候,不能忘记还有相当数量的“网络弃民”游离于数字世界之外,或者冲浪技术不够高超而遭遇不公,甚至网络“欺凌”。秦健教授对于数字公民现象进行了全面的阐释,并把数字公民所需具备的能力总结区分为九个方面,分别是:信息查找能力、电子商务能力、交流能力、数字素养、数字礼节、数字法律、数字权利与责任、数字健康与数字安全等,她特别指出传统社会中的相关做法和规则在数字世界可能完全失效,需要建立新的行为准则和道德规范,尤其要关注其中的一些负面影响。

公众科学和数字公民是当前数字学术发展的一种新现象、新趋势,图书馆不可能游离于整个数字学术大环境之外,它们与图书馆当前的发展有着密切

① 参见: <http://asn.jesandco.org/resources/D25275893>.

的关系。首先,图书馆已经开始利用众包方式来丰富和改善自己的资源系统,特别是一些数字人文系统;其次,越来越多的图书馆通过学科馆员、数据馆员等岗位直接参与到科研过程中去;第三,图书馆的资源管理门户向知识服务平台进化,本身提供了很多 2.0 交互功能和社会化功能,提供公众交互的可能;最后,图书馆作为一个公立机构能够对科研道德和学术规范起到一定的保护、正误、校勘和去讹作用,这其实也是图书馆长期以来作为一种规范控制机构的功能延伸。

秦健教授给大家展示了数字学术的完整图景^[17],她认为数字学术是利用数字化工具手段和数据进行知识生产和知识扩散的完整过程,伴随了学术研究生周期全过程。她列举了目前常用的数字学术工具,如 Zotero、Evernote、Dropbox、Diigo、Google Drive 和 Scoop. it 等,这些工具都具有很强的社交性,说明数字学术使学术的参与性大大增强,促进数字公民的增长,带来群众智慧。未来重大的科学进展有可能是成千上万人共同努力的结果,有学者笑谈,可能会有成千上万的数字公民因为参与了民众科学的项目,而同时获得诺贝尔奖。

无论如何,数字学术将是未来科学研究的一般形态,它具有对象的数据化、过程的网络化、组织的虚拟化、参与的公众化、成果的开放化、出版的语义化和交流的智慧化等特点。但它也不是没有负面效应,例如民科和反智现象等,这些现象只能在同样的数字学术环境中,通过加强交流、消除信息的不对称性等方式加以克服。

脸书和微信这样的平台正在发展成为人们须臾不可或缺的“社会机器”,人们在其中的创造成为生活的一部分,而科学研究也将如此。张晓林在报告中指出,未来的科学研究、内容出版,知识服务等将会由一种线上线下交织的“社会机器”来提供,人的活动只是其中的一部分,这是完全有可能的。

4 结语

数字图书馆虽已经历 20 多年的发展,但会议所反映出的趋势表明,其创造力远未停止。数据化和智慧化两个趋势并非本次论坛首次提出,但其最新进展和新的内涵在会上得到了充分诠释;数字学术和公众科学是本次论坛首次提出的主题,虽然其与图书馆的关系尚不十分明朗,但却是图书馆所依赖

的知识环境的重大变化,必然会影响到图书馆的管理与服务。这三方面的趋势是互相促进和紧密交织的:数据化是数字学术的重要原因,也是智慧化的前提和基础;公众科学是数字学术发展的一个民主化结果,顺应了网络化发展低门槛化和人人参与的潮流,其实每一项信息科技的发展都是这种趋势的助推器;智慧化是当今网络技术进步的一个总趋势。这三方面趋势正在带来数字图书馆新的变化甚至新生态,这也是本次会议的中英文主题:数字图书馆新生态和重塑数字图书馆(Reborn Digital Libraries)的用意所在。

管理学大师彼得·德鲁克(Peter F. Drucker)说过,“预测未来的最好方法就是创造它”。论坛的成功举办,说明数字图书馆这一开拓了 20 多年的领域并未过时,已召开了 13 届的数字图书馆前沿论坛也常开常新。图书馆行业新生事物依旧层出不穷,虽然一些人已产生审美疲劳,但“革命者”永远年轻。

参考文献

- 1 林海青. 从数字图书馆到数据图书馆——对学术图书馆发展的展望[EB/OL]. [2016-12-16]. http://www.sal.edu.cn/news_info.aspx?id=8892.
- 2 成鑫. 基于图书馆数据内容的精准化协同知识服务[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/基于图书馆数据内容的精准化协同知识服务_同方知网成鑫.pdf.
- 3 Tony H, Stewart T, Kristin T(Eds.). The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft Research, 2009: 177-183.
- 4 任树怀. 多语种数字学术服务平台的探索与实践[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/多语种数字学术服务平台的探索与实践_任树怀.pdf.
- 5 孙翌. 智能图书馆的应用尝试与设想[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/智能图书馆的应用尝试与设想_上海交通大学孙翌.pdf.
- 6 雷向欣. 数据挖掘分析对比之数字资源 VS 实体书[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/数据挖掘数字资源 VS 实体书_阿法迪雷向欣.pdf.
- 7 杨新涯. 数字资源管理+门户+服务联盟[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/数字资源管理_门户_服务联盟-重庆大学图书馆杨新涯.pdf.
- 8 林海青. 以关联数据实现规范控制:问题、方法与案例[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/Linked authority data_林海青.pdf.
- 9 曾蕾, 范炜. 《关联数据能力指标》及学习网站资源介绍[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/LD4PE-training_曾蕾_范炜.pdf.
- 10 LD4PE. Exploring linked data[EB/OL]. [2016-12-16]. <http://explore.dublincore.net/>.

- 11 范炜译. 关联数据能力指标[EB/OL]. [2016-12-16]. <http://explore.dublincore.net/wp-content/uploads/sites/2/2015/09/LD4PECompetencyIndex-chinese.pdf>.
- 12 刘炜. 机器智能时代的图书馆数据服务[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/机器智能时代的图书馆数据服务_刘炜.pdf.
- 13 CNNIC. 第38次中国互联网络发展状况统计报告(附报告全文)[EB/OL]. [2016-12-16]. <http://www.askci.com/news/hlw/20160805/17065450828.shtml>.
- 14 Clay Shirky 著; 胡泳, 哈丽丝译. 认知盈余[M]. 北京: 中国人民大学出版社, 2012: 1.
- 15 赵宇翔. 基于群体协作的公众科学: 数字人文的契机与挑战[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/基于群体协作的公众科学:数字人文的契机与挑战_南京理工大学经济管理学院信息管理系赵宇翔.pdf.
- 16 同 15.
- 17 秦健. 数字公民的理念与图书馆在数字时代的社会责任和服务[EB/OL]. [2016-12-16]. http://society.library.sh.cn/sites/default/files/digital_citizenship_and_Library's_Social_Responsibility_and_Services_in_the_Digital_Era_美国雪城大学信息研究学院秦健.pdf.

作者单位: 上海对外经贸大学图书馆, 上海, 201620

上海图书馆, 上海, 200031

收稿日期: 2016年12月19日

Digital Scholarship and Citizen Science: the New Environment to Reborn Digital Libraries

—Summary on the 13th Advanced Digital Library Seminar

Xie Rong Liu Wei

Abstract: The application of digital technologies in the library is developing further, presenting three new directions: from digitization to datalization; big data and cloud computing bringing intelligence and comprehensive crowdsourcing and citizen participation. At the 13th Advanced Digital Library Seminar, which was held in Shanghai Library, the theme Digital Scholarship and Citizen Sciences was first proposed in China. Along with the Datalization and Becoming Intelligent, these three trends are bringing the renaissance and new ecologies to the digital library.

Keywords: Digital library; Digital Scholarship; Citizen Science; Digital Humanity

《国际图书馆学期刊》征稿启事

《国际图书馆学期刊》(*The International Journal of Librarianship*)是由美国华人图书馆员协会(CA-LA)编辑出版的开放获取期刊,系图书馆学、信息科学领域的同行评审期刊,一年两期。作为具有国际视野的新创期刊,旨在分享跨界图书馆学新理念和新进展。在成功推出创刊号后,现征集特刊稿件。特刊将聚焦数据服务与数据管理这一迅速演变的领域,和全球图书馆员正在面对的层出不穷的新技术、新规则和理论上的新挑战。本刊邀请图书馆学实践者和研究者分享其经验和策略,欢迎任何关于学术数据服务、研究数据服务和数据图书馆学方面的论文。

包括但不限于以下主题:

- 定量和统计素养、数据引用及外延
- 数据关联、分析和可视化工具
- 期刊及资助政策
- 数字人文和定量数据
- 数据监管与保存、研究数据管理
- 编目、元数据标准和数据文档标准
- 馆藏发展、存取模式和开放数据
- 科学、工程、地理信息和大规模数据

请登录 <http://journal.calajol.org> 提交稿件,首次登录需注册,投稿截止日期为2017年5月1日。来稿要求约3000—5000字,采用美国心理学会(APA)格式。对高质量的论文可适当放宽长度要求。提交的论文必须是原创且从未在其他期刊或会议上发表。特刊将于2017年夏季出版。如有任何问题可咨询 Grace Liu(gliu@uwindsor.ca)。