



个性化知识服务的 实践与技术

Practice and Technology of
Personalized Knowledge Service

中国科学院上海生命科学信息中心 编
汤 江 主编



科学出版社

个性化知识服务的实践与技术

中国科学院上海生命科学信息中心 编

汤 江 主编

科学出版社

北 京

内 容 简 介

本书内容为第二届“图书馆个性化知识服务研究与实践”学术会议的论文集,围绕“个性化知识服务中的技术发展”的主题,就云计算、数据管理、语义搜索、自动标引、DOI 系统等信息领域前沿技术在个性化服务中的应用开展了广泛的研究和讨论。本书分为“技术篇”和“案例篇”两部分,集中展示了国家图书馆、中国科学院图书馆、上海图书馆、中国艺术研究院图书馆,以及华东理工大学、美国南佛罗里达大学等高校图书馆的研究成果和实践应用体验。本书可供图书馆信息服务人员和科技情报人员借鉴和参考。

图书在版编目(CIP)数据

个性化知识服务的实践与技术 / 汤江主编. —北京: 科学出版社, 2013. 1

ISBN 978-7-03-035997-1

I. 个… II. 汤… III. 图书馆工作—研究 IV. G25

中国版本图书馆 CIP 数据核字 (2012) 第 268147 号

责任编辑: 田慎鹏 许治军 / 责任校对: 纪振红

责任印制: 钱玉芬 / 封面设计: 耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2013 年 1 月第 一 版 开本: 787×1092 1/16

2013 年 1 月第一次印刷 印张: 12 1/2

字数: 296 000

定价: 48.00 元

(如有印装质量问题, 我社负责调换)

图书馆个性化知识服务的发展

（代序）

一、个性化知识服务的发展

图书馆个性化知识服务从服务理念、服务模式、服务技术、服务评估出发，研究用户需求、学科文献组织、信息评价、信息内容的提炼、信息定制和推送、文献引用记载、用户信息空间构建等文献情报方法。随着图书馆个性化知识服务的实践和推进，人们深刻认识到，计算技术和网络技术是当代图书馆存在的方式，是图书馆与读者，读者与读者互动的实现基础。图书馆个性化知识服务是基于或者说是依赖于计算技术和网络技术的，这是知识信息服务特性决定的，也是区别于其他行业个性化服务的不同的地方。分析图书馆个性化知识服务和用户使用的数据信息已成为重要发展方向。

就科技对人类文明进步的影响而论，计算技术的发展意义可以说是基本而深远。从1617年奈皮尔发明乘除器，到2010年苹果公司推出iPad平板电脑，我们的生活已因此而全然发生改变。那么2012年之后的10年、100年、1000年，计算技术将会给人类带来什么？

从现在看，计算技术和网络技术已创造了新的信息服务环境并深刻改变了人们的信息获取手段和方式方法，形成了新的信息生态环境和系统。从这点出发，图书馆个性化知识服务的理念和服务手段已经不单单是专业图书馆的发展方向，许多综合性图书馆和公共图书馆也越来越重视开展个性化知识服务业务，为读者提供便捷的、有针对性的、基于读者行为研究和分析的服务。这已经成为当今国际图书馆界关注和研究的热点。

二、一个程序化的世界对图书馆的挑战

我们处在一个程序化的世界里。数以亿计的处理器已经被嵌入智能型手机、轿车、电器等环境中。这些感应器发送和它们周遭环境有关的数据流，并且将数据传送至远程数据中心——谷歌、亚马逊、微软、雅虎和苹果等的“云端”。数量庞大的“云端”技术，能激活那些需要空间感知的技术服务。比如：谷歌地图从“云端”获得数据来抽样调查车内移动电话的位置和位移状况，绘制出实时交通拥堵图。智慧城市的建设使智能电网方便地测算我们家中的用电状况；成百上千、成千上万的人们上传地理标签至Flicker, Yelp, Face book以及Google Plus。有人预测，10年后，超级电脑（亿亿次级的计算机）将由数10亿个处理器构成，并且“云端”也很有可能增长到这个数量级别。此类计算技术加上用来存储所有收到的数据流的海量数据库，将使模拟未来现实世界的方方面面成为可能。在这场演变中，世界将逐渐变得程序化。

图书馆如何在这个程序化的世界里提供知识信息服务，如何通过海量的文献知识信息

数据和用户的使用行为信息数据流,来调查分析用户的使用习惯,研究使用热点,描绘使用地图和知识之间的相互关系,使图书馆拥有更高能效,这是图书馆面临的机遇和挑战。运用这些数据,计算机可以记录使用和服务的模式,同时将从感应器搜集来的数据与数以万计的人进行比对。计算机将可以发现那些可能导致学科变化的数据,既可以通过基于“云”技术的语音识别功能提供有关服务模式的个性化建议,还能避免无效服务和使用的发生,使得图书馆的服务方式更加亲民。

例如:IBM从2006年开始,每年都要拟一份“5-in-5”的报告,提出近五年内可能实现的五项尖端科技。每项预测乍听起来都有点不可思议,然而,在第一个五年结束的时候,IBM已经用科技成果证明了自身的前瞻性和行动力。比如,2006年,预测远程手术可能实现;2007年,预测我们的手机有可能成为电子钱包,用来付费购票等;曾经预测过的全息影像技术也已初见端倪……

今年,IBM公布最新的“五年计划,认为未来的五年,将可能在生物密码、人工能源、垃圾邮件‘变废为宝’、移动设备、读心术等实现五项尖端技术。”主要内容大致归纳如下:

(1) 生物密码,崭露头角。每个人都有独一无二的生物特征,借助它,也许会成为防止窃取身份的唯一密码。“多元素生物测定”,就是在现实生活中,视网膜扫描或语音识别将取代复杂的密码,成为访问信息或秘密地点的认证工具。当我们走到一个自动取款机面前,只要简单地说出姓名或者让镜头“扫视”就能安全轻松地取到钱。同样,还可以利用这种方式检查智能手机的话费余额。如果以上的情况成真,人们将不用大费周章地创建、牢记银行账户和社交网站的密码。

(2) 人工能源,造福生活。“世界能源展望”年度报告称:到2035年,全球电力需求将以2.2%的年速增长。这一数据,催发电网布局。授权给消费者、让其参与决策判断,使能源行业在进一步加速创新周期、制定有效方案上有所突破,建立可持续发展的能源基础建设。为了实现这一目标,就必须有兼具实用、环保的工具提供给消费者。消费动机离不开方便、舒适、低成本和数字化,需要合适的工具和信息服务大众,从而实现能源保护的目。人工能源还包括家庭,因为人们的运动、马桶的水流等所有动能都会提供能源。麻省理工学院的研究小组曾发布一份报告,提出“寄生功率收集”的概念,收集人体运动时所产生的能量。配合着广泛的活动范围、强烈的动作幅度,鞋底处安装的一个小型天线就能通过“寄生功率”为手机充电。

IBM的科学家已经开展了相关研究,他们正在探究如何利用海浪的能量进行发电;与瑞士苏黎世电力公司EKKZ共同开展了一个新的试点项目,让消费者能够方便地使用移动设备为电动汽车充电,同时监控其能源费用;进一步开发锂空气电池,从而达到电动汽车“充一次电,行800公里”的目标。

(3) 垃圾邮件“变废为宝”。IBM预测,智能机将成为一个解惑答疑、捕获信息的工具,即使在无人操控的情况下,也能独立完成。IBM正尝试开发这样一种新技术:通过自动筛选、信息分类,将数据整合到应用程序中,当这些程序遇到具体、新鲜的相关信息时,会及时发送给用户。因此,当智能手机再次接到电子邮件、广告时,我们将不再感觉是一堆垃圾信息,取而代之的是带来“变废为宝”的体验。

搜索引擎已经给人们带来很大的便利。未来,将通过信息来寻找信息,从而发现更具价值、更有新意的东西。这对智能机的升级换代提出了新的要求。最终,人类与智能手机建立起足够的“信任关系”,从此不再需要一般性的购买,只要指定预算和所需服务,智能手机就会自动带来符合要求的高品质项目。

(4) 移动设备,终结数字鸿沟。移动通讯行业,悄然发生着转变:移动设备的价格逐渐“亲民”、购买较为方便、功能更加多元。如今,半数世界人口拥有手机,四分之一使用互联网。5年前,电话被当成语音装置,而不是像今天一样,作为一个带有数据、信息应用程序的工具。手机、电脑等满足快速访问需求的产品,通过自身优质特性,更富有生命力和竞争力了。移动电话提供了“足不出户”的教育、卫生、新闻信息以及其他服务。未来五年,数字鸿沟或将不复存在。社会连为一体:政府与市民进行有效沟通、医生对病人展开远程诊断、身处两地的研究人员举行虚拟学术会议……这一切,都源于这个智能、交融的数字网络世界。

(5) 读心术,不再是科幻。生物信息学专家已经设计出一款带有先进传感器的耳机,能够“先知先觉”人类将要做出的面部表情、行为意向。

这款“Emotive Life Sciences”耳机,看上去像一个挥舞着触角、盘坐在脑袋上的章鱼,能够“读取”大脑意图。当我们看到电脑屏幕上的立方体时,只要想一下“向左移动”,它就能自动做出相应行为。如今,用“意念”操控电脑屏幕上的文件不再是难事。科学家表示,未来五年,这项技术将首先在游戏、娱乐和医疗保健方面“投石问路”。医生可通过该技术掌握大脑活动方式,帮助康复治疗,甚至有利于了解自闭症等成因机理。IBM的科学家与合作伙伴正在开发一种软件,希望将Emotive Life Sciences”这款耳机连接到电脑、智能手机等设备上,如此一来,只要想一下呼叫对象,手机便会自动联系,真正实现“心想事成”。

从IBM预测我们可以看到,这一切的实现将对现在图书馆的存在方式和服务体系提出严峻的挑战,如何变挑战为机遇,是我们必须要面对和回答的问题。

三、一个创造阶段可能迈进的新领域

由互联网所代表的颠覆性技术会冲击旧有信息服务体系和方式,从而会诞生新的服务体系和服务方式,在这一体系中个性化知识服务的方式是不可或缺的,我们现在处于一个创造阶段。没人能想象得出互联网有多广袤。身处无法设计或预期的世界,我们只能紧跟技术的发展,不断迈进新领域。

一方面,我们需要几年前还很难想象的巨大的免费电子邮箱、电话、地图、大的图片空间;另一方面,别的人或机器阅读了你的邮件,那他们会寻找新的途径来控制你的将来。这样,有些事情会越来越复杂,有的是给人类社会带来欣喜的,而有些东西则是给人类社会带来威胁的,因此,我们要不断探讨孰轻孰重的问题。比如,Facebook和谷歌是一种快乐的交流方式呢,还是对我们隐私的一种威胁?

科技的发展还将深刻改变人们获取知识和接受教育的方式。

首先,将改变教育。

我们的教育方法从文艺复兴到现在一直没有多大的改变：从中学到大学，大部分教学依旧在坐满学生的教室里进行，认真听讲的学生不多，效率不高。直至今日，个性化教学仍很难实现。但本人认为，网络技术也许会创造一条成功之路。

现在免费精品公开课异军突起，已经成为网络教育的有效手段。据美国斯坦福大学统计，该校上传了3门计算机科学课程。在前1个月里，就有30万学生注册观看，观看次数达到几百万，并提交了数以千计的网络作业。从这些成功案例中，我们可以发现个性化在知识教育领域有广阔的发展前景。美国学者归纳总结后认为：首先，这些是专门为学生设计准备的视频课程，同时教师制作视频也很简单。其次，以短小的视频呈现学习内容取代持续1小时的教学课程，更符合学生的注意力持续时间。另外，具有较强学习能力的学生也可以超前学习，避免厌倦以及松懈。教师和学生用较少的准备就能运用较多的资料，并且不会因时间不足感觉到不安。

网络教育能够实现两个目标。一是学生能够接触到优秀的教师，在相同或更低的成本条件下，交叉学习能够带来更好的学习效果。二是对成千上万个国内或者国外的不能接受良好教育的学生来说，网络学习为他们打开了一扇窗户。

网络技术能够追踪到每一次点击：记录下哪一门课学生看了超过一次，他们在哪里中断，他们出了什么错误等。这些数据对理解学习过程，确定最好的学习策略来说是无价的资源。

据统计，2010年有关教育部门通过45项研究得出的一项分析表明，在线教育与当面教学一样有效，并且两者交叉的学习方法被认为比这两者都有效率。

其次，将改变图书馆服务方式。

在新的信息生态系统和学科间交叉协作模式下，专业图书馆的服务内涵与馆员的角色发生了许多重要变化，图书馆不能被动地等待研究人员来寻求信息服务，这样图书馆将边缘化，将失去知识信息服务中心的地位。图书馆要主动接触研究人员，并参与他们的研究活动，在数字图书馆舞台上扮演更加多元化的角色。美国南佛罗里达大学信息学院助教黄洪博士研究撰写的《美国学科图书馆对数据质量控制、科学家协作及数字监管的信息服务尝试》，列举和剖析了美国专业图书馆关于了解科研人员方的信息需求服务和科研协作支撑的实例，以及提供数字监控服务的具体做法与经验。这些案例不仅对专业图书馆，而且对于公共图书馆、大学图书馆也具有积极的借鉴意义。

近年来，图书馆个性化知识服务的研究和实践，产出了很多成果和鲜活的案例。把这些成果和案例介绍给有关学界，希望能对图书馆知识信息服务提供一些有益的借鉴，也正是编著出版这本书的目的。

知识信息服务通常是借助技术手段来实现的，但是很少像当代计算技术和网络技术与知识信息服务结合引发如此强烈的空前演变。当代科技正在将信息服务带入一个新的时代。当代科技为信息服务提供的手段与可能性直接地决定了信息服务的方式，而服务方式又推动着科技的发展与进步。许多曾经在科幻境界中被津津乐道的神奇般的镜像，正在被当代科技变为现实。图书馆界要充分认识和准备好迎接这个变化，这是极其重要的。

汤江 研究馆员

2012年3月30日

（上海市重点领域个性化知识服务体系建设与完善项目负责人）

目 录

图书馆个性化知识服务的发展(代序)	i
-------------------------	---

技 术 篇

美国学科图书馆对数据质量控制、科学家协作以及数字监管的信息服务尝试	3
云计算在个性化服务中应用的技术预测	9
信息供给、知识引导:图书馆服务中小企业过程中的交互融合与技术实现	18
Web 2.0 技术在图书馆个性化知识服务中的应用	22
基于关键词的知识点自动标注研究	28
DOI 系统在个性化知识服务中的应用探讨	33
面向生命科学领域的语义搜索引擎探析	42
数字图书馆个性化信息服务浅析	48
国内外图书馆微博客的应用比较分析	53
图书馆数字资源服务模式探讨	61
运用知识组织改进参考咨询知识库建设研究	65
DOI 在个性化知识服务系统的应用探讨	69
浅谈高校图书馆 Human Library 服务中的知识管理	75
上海研发公共服务平台个性化文献订阅服务系统技术分析与展望	80

案 例 篇

基于自存储的机构知识库建设研究及案例	91
产业引导型政府信息服务中的若干知识结构问题初探——公共管理理论的应用与实践	101
面向“重大新药创制”科技专项的知识组织和知识服务——中科院上海药物研究所图 情室的实践探索	106
知识组织在小型自助化阅览室中的应用——以上海图书馆自助阅览室为例的探索	112

OPAC 2.0 与“辨章学术,考镜源流”传统的新生	118
公共图书馆个性化服务推广策略浅谈——以厦门市图书馆为例	125
在知识管理态势下打造专业图书馆核心竞争力——以中国艺术研究院图书馆为例	131
个性化知识服务在艺术类图书馆中的探索和实践	136
图书馆网上参考咨询与其他相关业务关系之定量比较研究——以上海图书馆网上 联合知识导航站为例	140
一种新的基于 SaaS 服务的中小图书馆学科服务平台——DatabaseKD	145
基于知识服务的《申报》电影信息库构想	149
CBA 数据库个性化知识服务的设想	154
影响用户使用个性化文献订阅系统的若干因素及思考	158
中科院研究所图书馆个性化服务探讨	163
附录:国内外个性化知识服务平台一览表	167

技
术
篇

美国学科图书馆对数据质量控制、 科学家协作以及数字监管的信息服务尝试

Providing the Support of Data Quality Assurance, Researchcollaboration, and Data Curation in the U. S Academic Libraries

黄 洪*

摘 要 在新的信息生态系统和学科间高度交叉协作模式下, 学科图书馆的服务内涵与馆员的新角色发生了许多重要改变。图书馆不能被动地等待研究人员寻求信息服务, 而应积极主动地接触他们, 并参与他们的研究活动。学科图书馆在数字的舞台上扮演的角色更加多元化, 例如, 网络基础设施建设、协作研究的支撑、数字监管和典藏服务。学科馆员的角色也将更加多样化: 数据的保存以及监管、研究成果的发布与传播、合作网络的构建, 以及协助研究人员的协作。本文列举并剖析一些美国学科图书馆关于了解科研人员的信息需求服务和科研协作支撑的实例, 以及提供数字监管服务的具体做法与经验。这些案例的提出以及经验不仅对于学科图书馆, 而且对公共图书馆、大学图书馆也具有积极的借鉴意义。

关键词 数据质量 科学家协作 数字化监管 信息服务

在移动通信、网络、移动存储高度发达的信息时代, 图书馆的影响力是不断扩大, 还是在变小? 应该如何应对用户新的信息需求? 新型的图书馆信息服务模式包括手机图书馆、个人(自助)图书馆和网络图书馆, 使图书馆由传统的阵地服务朝泛在化服务发展。图书馆阵地服务强调物理空间、公共空间和学习空间的定制, 强调基本信息的获取。图书馆泛在化服务实质是阵地服务的时空延伸、信息细化及个性化, 以及强调信息的专题、定制和特性。图书馆泛在化服务需要信息技术的支撑。它是技术发展的结晶, 是以人为本的服务集成。在泛在化服务中, 用户参与程度颇高, 使得对馆员的要求很高, 对图书馆员、资料员、信息员提出了新的挑战。

图书馆已不能指望科研工作者和学者主动来获取帮助, 无论他们在哪里, 都必须去寻找能够获得利用信息的新途径。研究和学习活动在科研机构内部与机构间越来越具有合作性、跨学科性、国际性和虚拟性, 因此学科图书馆工作人员的责任更艰巨了。他们往往能起到协调员的作用, 对不同的生物情报源, 进行信息集合, 指导信息索引, 元数据标准的制定, 对学科发展趋势进行预测。他们要对新兴的数字图书馆, 数字档案馆和资料库提供

* 黄洪, 男, 1973 年生, 美国南佛罗里达大学图书馆情报学院, 助理教授, huanghon2003@gmail.com

理论和技术的支持,保证系统的工作持续性。他们要为不同的用户群体提供服务,为数字资源的长期保存、数字汇聚、信息架构、数据建模和数据标准进行指导,以提供可持续发展的基础。他们还要促进跨学科和多维度的交叉科学的基础设施建设和服务并根据需要提供专业的图书情报知识。我们以下讨论图书馆应当如何在这个新环境中调整它所提供的服务,以保持活力。

1 了解信息质量需求与协作模式的研究

近年来,学科间高度交叉与渗透,让研究人员不得不重新思考研究协作的方式,并提出更高的数据管理质量的要求。以生命科学为例,生命科学经过多年的发展,现在已形成一个高度汇集的交叉学科;同时也为图书情报科学研究提出了新的要求。生物情报学作为一门新兴学科,通过对生物数据的获取、存贮、检索、分析与预测,使大量分散和复杂的生命科学信息情报得以更好地加工与利用,及时转化为有利于社会的知识,从而为生物学家提供个性化信息服务,使之进行更高效的科研活动。我们可通过以下两个例子,了解学科馆员如何提供高度交叉性和协作性的专业服务。

1.1 了解科学家的数据质量的信息需求

为什么信息专业人员在合作研究中起到重要作用?这是因为他们能了解不同学术背景、不同科学家角色变换时的特别需要。随着基因组海量数据的急剧增加,多种多样的相关基因组注释工具迅速建立起来。这就要求进行高质量的基因组数据注释,充分优化原有的注释资源和工具,确定正确的注释策略,例如,在医药信息方面,使注释的数据更好地推动转化医学以及临床和科研实践的广泛应用。这样的研究有助于更好地了解科学家的看法和优先考虑的基因组数据注释质量的要求和注释技能需要。

此外,由于基因组注释时扮演的角色及其教育背景的差异,一般用户和基因数据实际注释者可能侧重于不同的数据质量方面有关的基因组注释要求。由于我们缺乏对不同用户和学术背景的基因组注释研究人员需求差异的系统了解,就很难对现有的基因组数据注释质量提供保证。研究并了解上述的不同用户对注释质量的差距,确定不同用户对数据注释质量以及技能的要求。提出指导基因组注释的标准流程,这为质量评估和基因组注释数据管理提供了有效方法和途径。

基因组注释质量的研究是利用数据质量管理的总体框架对基因组数据的质量和分类的数据注释技能进行评估。在基因组注释研究的框架下,为了解注释角色不同质量要求的差异,将受问卷调查的一百五十名基因组研究专家按基因注释角色分为三组(普通用户、数据注释人员以及双重角色的用户)。同样,为了解学术背景不同的质量要求的差异,我们将问卷调查的专家根据学术背景分成三类,即生物学背景、生物信息学背景和双重背景的专家。针对数据质量和基因组注释技能的质量需求的差异进行了探讨,以确定质量概念、特定的注释角色或教育背景的个人用户,在基因组注释过程中考虑的重要标准。分析结果显示,十七个数据质量维度可以归纳为五个数据质量要素,而相关十七个注释技能维度归

纳为四个要素。调查结果分析显示不同角色和教育背景的用户,对基因组注释的质量有不同的要求。不同角色的用户看重不同的数据质量维度和技能不同。一般用户间接依靠来源可信度评估的注释记录的数据质量。作为数据注释实际操作人员,因为直接评估数据质量,则更多关注基因组注释的可用性和注释的完整性。

因为用户最终决定注释信息的实用性和价值,用户之间需要更紧密的合作,并综合考虑不同的数据质量的需要,对基因注释过程进行更好的质量控制。注释质量的评估管理框架的提出,有助于最大限度地提高注释的效率和准确度。由此产生的质量评估框架可以为学科馆员对基因组数据资源注释的政策制定和设计数据质量监测提供研究战略。这项研究的结果也可作为研究人员提供数据质量控制和注释技能的培训课程。

1.2 科学家协作模式的研究

在交叉学科发展迅速的今天,学科馆员如果了解科学家协作的模式,便可以开发如何更高效转化科学成果的途径,更好地提供相应的信息服务。例如,佛罗里达州立大学的情报学院开展了一项美国国家自然科学基金委员会(National Science Foundation)资助的研究科学家协作的研究项目,了解科学家协作生命周期的形成和对长期的科研合作模式的探索。这项研究是立足于相关的信息小世界(Small World)的理论概念的社会和组织因素分析来研究社会规范、社会类型、社会价值和信息的行为。该项目研究提出以下研究问题:一个研究单位的社会和组织是如何构成的?如何更好地从支持短期实验为重点的多学科研究向长期的生产和创新科研计划的科研合作过渡?该研究的目标是开发和验证一个或多个生命周期科研协作模型,支持短期合作向长期合作过渡,认识在离散的研究地区以实验项目为主的交叉学科科研团队的协作模式,从而推动创新和提高科研成果的转化。具体来讲,该研究围绕美国国家高磁场实验中心(National MAG Lab)的组织结构进行调研。该中心是一个跨学科的多地区(佛罗里达州塔拉哈西,新墨西哥州洛斯阿拉莫斯国家实验室,佛罗里达州盖恩斯维尔)交叉学科研究中心。通过对研究人员访谈,问卷调查,研究他们的文章发表的引用特征,勾画出该单位科研人员的协作知识关系图,建立数据监管的生命周期模型。

2 数字研究基础设施和数字监管的研究

美国国家自然科学基金委员会(National Science Foundation)和美国国立卫生部(National Institute of Health)要求所有获得资助的基金都必须把研究成果数据共享化。科学家在做好研究的同时,也要考虑到如何让他们研究成果更好地与社会共享。为满足这方面的要求,科学家们就需要图书馆员的协助其进行科学成果的数字化整理,数字保存,数字监管和数字信息发布。另外,日益数字化和越来越多的数据密集型的科学和教育科研活动,急需对数字整合提出更高的管理要求。因为数据不仅是现有的研究成果,随着新的假说的提出,研究资源的投入,会衍生出新的科学见解以推动科技创新。以下两个例子说明了在数字化基础设施和研究协作网络建设的学科图书馆所做的工作。

2.1 DataNet

面对这一信息时代的海量数据集和数据流的复杂性挑战,如何开发出新的方法、管理结构和技术的多样性与灵活性,需要通过创建一个国家典范的和全球数据研究基础设施的机构(DataNet)为不同地区的研究人员,提供数字管理的方便,以促进他们的科研活动和学习。

DataNet 提供了新类型的图书馆个性化信息服务,其目的是将整合相关的学科、图书馆和档案馆网络基础设施、计算机和信息科学以及不同领域的科学专业知识。DataNet 可提供可靠的数字化保存、访问、整合和数据分析能力。为满足对数据长期管理的跟踪需要,DataNet 计划对数十年之久的科学研究数据进行保存和监管,并在此基础上运用信息分析与预测的方法,进行持续的预测,为决策提供依据,以适应技术的变化和用户的需求与期望。

DataNet 的建立也推动了计算机和信息科学和网络基础设施的研究和开发。DataNet 作为一个可互动操作的数字保存和接入网络的虚拟组成要素,通过视觉展示,确定对个性化用户的最佳方案,并结合前沿研究,建立长期的技术与经济可持续发展的可行模式。可以成为科研群体所在区域作为理性的投资,在数字资源长期保存和访问的基础上,构建成为一个强大的和有弹性的国家和全球的数字数据框架。

DataNet 还提高了研究人员的合作性、跨学科性、国际性和虚拟性,具体表现在以下几个方面:①信息可视化。满足需求的关键数据,挖掘新的重要发现,提供创新和学习的机会和能力,进行科学、工程研究和教育的提高,并引导相关机构和组织的长期可持续性发展。②组织结构管理。提供全方位的专业知识和网络基础设施能力,确保积极参与和配合个人、组织和部门的广泛多样性的有效利用;并作为数字资源长期保存和网络服务能力的合作伙伴,确保有效的管理和领导。③提供完整的数据生命周期的管理。促进资源和对象的研究。另外,DataNet 促进新的学习工具的开发,整合了各级科研和教育的活动,也积极推动了社会投入和合作伙伴的参与,并及时提出全面的评估和评价方案。

2.2 研究社区内容网络与服务(VIVO)

美国高校图书馆开发出来的 VIVO 研究网络(<http://vivo.ufl.edu>)是一个开放源码的语义网(Semantic Web)应用程序设计。它的建立可以推动研究开发和研究人员之间的协作。跟别的社区网络(Social Network)不同的是 VIVO 尽可能地收集了权威性的情报源。因此 VIVO 及其相关系统的数据信息来源很多。可以来自权威机构的记录与报表,科研基金数据库,出版物数据库和研究人员报表系统;提供一个更加一体化的框架,在研究人员参与、资源和工具配置方面,减少了投入,并提高解决复杂问题的分析能力和加强多学科的研究。

由此产生的相互联系,以及与其他网络的可视化工具,可让研究人员发现新的内容,以利于找出专家和潜在的跨学科合作者。此外,学生可以利用该社区而找到相应的导师和

感兴趣的科研项目。VIVO 还提供管理员的权限,以便学校的院长或副总裁了解最新的科研活动动态,从而更好地进行管理。VIVO 还通过研究协作的知识图谱来展示具体研究方案和管理机构的数据。

学科馆员应认真考虑参与机构网络基础设施的发展建设,重新调整学科图书馆的使命,并重新定位图书馆作为一个学术研究的全面合作伙伴。馆员的信息组织、管理经验和学科专业知识,使他们成为在执行研究网络的理想合作伙伴。特别是在发展的网络内容的管理、定位和选择的主题词库,发展以用户为中心的界面设计,为潜在用户进行培训,为项目管理提供协助等方面发挥作用。

通过 VIVO 这一社区内容网络工具,能管理大的跨学科的团队,提供信息技术的支撑,赋予团队精神和团队动力,跟踪项目性质的变化,平衡工作量,从事和参与更广泛的社区项目管理,使图书馆扩展其传统的角色,并获得新的合作伙伴和技术知识。VIVO 项目成员能够将这些经验推广到如 e-Science 和转化科学项目管理中去,并分享这个知识管理的工具。

2.3 为科学家提供数字保存与监管的服务

随着图书馆服务向泛在化服务的延伸,研究图书馆需要被重新定位。作为活跃的知识部门,它们能够延伸到所在的整个服务机构,不管数字内容在哪里创建和维护,都能得到管理者的专业指导。数字监管(Digital Curation)指的是人们在数字信息的使用寿命内维护并提高其价值所采取的行动,包括创建数字内容的过程。它主要指“为尽可能保证数字材料可持续访问所必需的一系列受控活动”。不同的学科领域对数字保存的要求是有差异的。首先,数据存储问题很复杂:如数据有什么变化,何时以及如何变化。其次,收集的数据方式非常多样,如果基于网络的内容,可能需要重新建立数据存储和查询功能,因而有额外的信息服务要求,以满足不同的主要用户群体和再利用群体的需求。我们如何才能增加数据的使用价值,访问和使用量呢?

佛罗里达州立大学图书馆在为研究科学家数字存储和数字监管服务方面作出了有意义的尝试。例如,在图书馆建立数字监管的 Diatomscapes 档案数字化馆藏。这个项目是对生命科学数字对象在数字生命周期内进行监管的应用案例研究。该 Diatomscapes 数字馆藏通过 JEOL JSM-840/FEI 纳米扫描电镜,获得二氧化硅的三十八个图像扫描的电子 microscopes 创建一个电子档案。与美国佛罗里达州立大学的生物科学家 AKSK Prasad 博士密切合作,学科馆员对生物硅图像进程数字监管。首先,确定数字的生命周期模型和保存的存储的应用模式,与佛罗里达州立大学特藏部密切合作,通过 MetaArchive 保存档案 TIFF 图像(DigiTool)的数字馆藏技术,提供数字访问,使用和重用的技术支持。该项目的目标是提高公众对研究的认识,针对生命科学数字对象的生命周期模型的应用开发,并推广生命科学未来跨学科的研究与应用领域的可扩展数字资源长期保存策略,通过社交网络工具和 OCLC WorldCat 对数字典藏的访问和使用进行推广。

数据监管涉及了数据管理的多个方面。它需要了解数据来源,对数据资源进行选择与评价,元数据标准和协议的制定,提出数据收集方法,了解数据生命周期的学科特性,对

基础网络设施及知识库进行构建,理顺工作流程,以更高的支持存档和共享。而且,数据监管对知识产权问题数据再利用的价值、研究人员合作政策、图书馆员的角色和技能提供了有意义的实践经验。

总结

图书情报人员应该主动参与科学家的研究活动。数字化信息服务环境提出了更高的信息服务挑战,学科馆员要了解复杂的跨领域研究文化差异,直接参与科研作活动,并作出直接贡献。学科馆员应在数字环境下提供跨学科研究活动的人员之间的高度协作,并适应不断变化的用户需求。某个领域的专家可能没有像学科馆员那样的广阔知识面,缺乏对数据处理技能,以及面对处理信息挑战的能力(例如,“Information Overloading: 信息超载”)。因此,专家学者也需要相关的信息素质的培训。学科馆员应该专注于数字数据的收集和管理,开发和设立相关的培训课程,不断实习,发扬共享数据监管的专长,与科学家紧密合作,共同推进交叉学科的研究和数据监管的发展。

参考文献

- Beagrie, N. (2006). Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation*, 1, 3-16.
- Garcia-Milian, R., Norton, H. F., Auten, B., Buhler, A., Davis, V. I., Ferree, N., Holmes, K. L., Johnson, M., Tennant, M. R., Conlon, M. VIVO Collaboration Role of Librarians in the Development of Computer-Mediated Social Networks: Challenges and Lessons Learned From VIVO Implementation and Outreach. Biomedical and Life Sciences Division Contributed Papers, Special Libraries Association, 2011, Philadelphia, PA.
- Huang H., Stvilia B., Jørgensen C., Bass H. (2012). Prioritization of data quality dimensions and skills requirements in genome annotation work. *Journal of the American Society of Information Science and Technology*, 63 (1), 195-207.
- Plato L. Smith II, (2008). Where IR you? Using “open access” to extend the reach and richness of faculty research within a university. *OCLC Systems & Services*, 24 (3), 174-184.