

大数据视阈下我国数据人才培养的思考*

马海群¹, 蒲攀²

(1. 黑龙江大学信息资源管理研究中心, 哈尔滨 150080; 2. 黑龙江大学信息管理学院, 哈尔滨 150080)

摘要: 对大数据时代国内外数据人才需求现状进行分析, 并将数据人才分为数据技术人才、数据管理人才、数据安全人才、数据分析人才、数据政策人才、数据开放人才和数据科学家这七种类型。在对这七种数据人才进行深入剖析的基础上, 提出我国数据人才培养的三种路径, 最后指出图书情报学科对数据人才培养的定位与前瞻。

关键词: 数据人才; 人才培养; 图书情报学科

中图分类号: G259

DOI: 10.3772/j.issn.1673-2286.2016.1.001

早在20世纪80年代, 美国就有人提出“大数据”的概念; 2004年社交媒体产生之后, 数据大爆炸全面来临。大数据时代的到来, 催生了全新的思维变革与科学发现, 我们不再纠结于因果关系, 因为“全数据”分析为我们展示了所有既定的事实, 假设似乎再无必要, 原因也似乎不再重要。微软研究院首席研究员凯特·克劳福德在《对大数据的再思考》中对大数据提出的质疑却引人深思: 对大数据“过度解读”的倾向, 会导致分析结论的不切实际; 对数据的处理与应用不当, 会造成隐私泄露^[1]。Simón等认为, 公共信息是社会变革的动力(如气象、地理、交通信息, 以及图书馆、档案馆、博物馆所保存和管理的信息), 能够提供数字化的经济业务服务, 应该面向公众开放。但是目前还不清楚如何将大量的公共数据引入公民的教育、卫生、文化等领域来提供普遍服务。他指出, 这时候就需要一个新的角色来承担这项历来由图书馆承担的任务, 即信息专业人士^[2]。因此, 在“死”的数据与“活”的应用之间架起桥梁的信息专业人士显得至关重要, 将直接影响大数据的终极产出, 我们将这类人员称为“数据人才”, 它既包括被美国《哈佛商业评论》杂志列为“21世纪最性感的职业”的数据科学家, 也包括整个数据处理生命周期中涉及的全部数据业务人员。

众所周知, 21世纪最缺的就是人才。在大数据环境下, 我们需要在人才前面加上“数据”属性。近年来, 为了与大数据的发展保持一致, 世界各国纷纷开始数据人才的培养。大数据技术涉及应用数学、统计学、人工智能、可视化、软件工程和管理科学等多领域的专业知识, 人才培育复杂, 在我国尚处于起步阶段。2015年9月6日, 由上海市数据科学重点实验室主办的“首届数据科学家大会”在上海召开, 大会以“WHY数据科学家”为主题, 讨论数据科学理论、大数据技术、大数据人才挑战、数据科学家培养等问题, 试图促进数据科学发展、大数据技术研究与应用、数据科学人才培养^[3]。

1 大数据时代对数据人才的需求现状分析

20世纪90年代, 计算机工程师和金融工程师是最炙手可热的职业。而今天, 数据科学家是猎头公司的抢手货。《华尔街日报》2014年报道称, 3年前数据科学家这个职业头衔基本还不存在, 如今已成为高科技劳动力市场上最热门的职业之一^[4]。根据麦肯锡研究院2011年发布的报告, 到2018年, 全美范围内对具备量化分析等分析技能的“大数据”相关人才的需求, 将会达到

* 本研究得到国家自然科学基金重点项目“开放数据与数据安全的政策协同研究”(编号: 15ATQ008)和黑龙江大学研究生创新科研项目“大数据环境下我国开放数据政策模型构建研究”(编号: YJSCX2015-066HLJU)资助。

400万之巨。此外,在大数据相关的管理与分析人员方面,还有可能出现150万的用工缺口^[5]。埃森哲的调查显示,美国新增数据科学家职位的数量将占全球新增总量的44%,但美国只能供应23%的人才,有近3.2万人的缺口。到2018年,仅美国和英国,数据科学家职位的增长速度将是其他职位的五倍,是金融服务等信息密集型行业职位的四倍^[6]。根据一家负责数据类相关职业招聘的机构开展的调查:数据分析行业的从业人员中,89%的人基本上每月都会通过领英网(LinkedIn)等网站拿到新的邀约,25%的人被其他企业联系的频率更是高达每周一次^[5]。CrowdFlower对全国的数据科学家开展的调查显示,近80%的数据科学家认为,没有足够的数据科学家在这一领域工作,有23.5%的数据科学家希望聘用更多的数据科学家^[7]。2014年2月,Gartner发布报告称,截至年底,17%的全球大型企业将会在管理层增加首席数据官(CDO)职位^[6]。

据业界专家估算,未来五年,中国大数据人才需求至少为100万人,而目前已有的大数据人才尚不足10万^[8]。一个拥有博士学位的数据科学家的起薪通常是六位数,工作两年后,就可以轻松赚到20万至30万美元的年薪^[4]。我国大数据产业的领跑者——贵州省,近年来已经意识到数据人才对于“云上贵州”发展的推动与制约,并积极采取与大数据企业联合培养、给予引进人才优惠政策等方式开展人才培养计划。但算上人才培养投入和产出的周期,有企业家分析,贵阳大数据人才的“荒漠期”将会持续两三年^[9]。可见,当今数据人才已成为国际社会争夺的重要人力资源,而相关数据人才的短缺是制约我国大数据、乃至开放数据发展的主要瓶颈。

在我国,目前对数据人才需求最多的行业主要集中在金融、通信、网络等领域。国务院2010年印发的《国家中长期人才发展规划纲要(2010-2020年)》(以下简称《纲要》)将突出培养造就创新型科技人才、大力开发经济社会发展重点领域急需紧缺专门人才作为我国人才队伍建设的主要任务^[10]。数据人才既属于创新型科技人才,又是当下经济社会发展重点领域急需紧缺的专门人才,理应在《纲要》的指导下,广泛联合全社会各组织机构的力量,大力进行人才需求计划的制定、人才优惠政策的保障、人才培养与发展机制的建立。

2 大数据时代数据人才的类型分析

在传统的大学培养和教育体制下,我们已经在相

关专业学科领域储备了一些数据人才,包括统计学、数学、计算机科学、管理学、经济学、社会科学、人工智能、可视化等领域。但在大数据的环境下,Simón等研究强调,信息专业人员必须掌握新的技能,使他们能够承担信息管理的新角色,即数据管理(开放数据)和内容管理(开放内容)^[2]。因此,传统专一学科的人才已经很难适应社会发展的需求,倘若在大数据处理的整个过程中只是这些专一人才“硬性”的对接,那么必然导致大数据价值产出的“断截”,如此造成包括人力、数据、资本、技术等社会资源极大的浪费,这是我们不愿见到的。在大数据时代,我们需要数据的复合型人才,即能够将扎实的专业知识背景熟练运用到所在行业和业务领域中去,实现数据处理的“无缝连接”。笔者认为,大数据时代数据人才主要分为以下几种类型。

2.1 数据(处理)技术人才

这类人才主要从统计学、信息技术、软件工程领域诞生,主要负责数据处理的全过程,即数据的获取、存储、清洗、加工、建模、传输和诠释,数据架构师、数据工程师都属于数据技术人才。大数据区别于传统数据的首要特征就是Volume,即数据的体量大,现今全球的数据总量已经从TB级别跃升到PB级别乃至EB级别,单从这一点来看,数据的处理已经不是传统技术手段可以完成的。大数据处理遵循“1秒定律”(或称秒级定律),一般要求在秒级时间范围内给出数据分析结果,这个速度要求是大数据处理技术和传统的数据挖掘技术最大的区别^[11]。数据体量的跃升、秒级定律的要求,都对新时期数据处理技术提出了更高更大更难的需求。

在小数据时代,传统的数据获取大多采用人工的数据收集,例如问卷调查、人口普查等,由于当时的数据量远不如现在,所有人工的方法尚可。自从2004年社交媒体产生之后,数据大爆炸全面来临,“信息社会”也终于名副其实,每秒有数万TB的新数据产生,其中大部分来自社交网络与搜索引擎的贡献。数据总量的急速膨胀,使得传统的数据处理模式已经无法适用。在以关联数据和非关联数据为主的大数据时代,在传统关系型数据之外我们有了以Hadoop为代表的HBase、以MongoDB为代表的NoSQL、以MySQL为代表的开源数据库^[12],也拥有了以数据库方式工作的免费开源的数据清洗工具OpenRefine、DataWrangler和Google Refine等。但类似这些数据处理技术与工具的掌握,只

在一小部分IT精英中,这部分人才的数量还远不能满足如今大数据对各行各业的渗透。因此,要实现大数据处理,相关数据技术人才的发现和培养是关键。

2.2 数据管理人才

这类人才主要从计算机科学、管理学、经济学领域诞生,主要负责对数据的保存、管理、维护和运营。各个行业都有管理人员,大数据也不例外,只不过大数据时代的管理人员所面临的更多是“数据”这个特殊的管理对象,而能够适应这个特殊性的管理人员,才可能成为数据管理人才。

数据管理在经历了人工管理、文件系统和数据库系统三个阶段后,如今实现了数据的长期保存、数据共享和独立性,也已出现了统一管理数据的专门软件系统——数据库管理系统^[13]。但我们所指的数据管理人才不仅是能实现对数据需求管理、元数据管理、数据质量管理等方面的传统或面向应用的数据管理,还必须具备数据维护和运营的能力,即面向业务的数据管理。优秀的管理人才,应该能够敏锐地捕捉所管理数据的核心价值,能够通过业务流程发现数据增值的空间,能够在“数据”与“价值效益”之间找到契合点,能够及时准确判断数据的折旧值,能够合理利用数据废气实现再增值。

2.3 数据安全人才

这类人才主要从政策科学、计算机科学、社会学、伦理学领域诞生,主要负责对数据安全(包括数据本身和数据防护安全)的维护和保障,包括维护数据隐私、防止数据盗用和滥用、保护加密数据、阻止黑客攻击、建立数据安全防护体系等。

2015年7月1日,以“数据驱动安全”为主题的2015中国互联网安全大会(ISC2015)在北京召开,约2.5万名安全行业人士共同探讨智能移动终端攻防、网络安全空间战略、网络空间安全法律、大数据安全等话题^[14],该大会2014年的主题为“互联世界,安全第一”。可见,网络安全、信息安全乃至数据安全在我国越来越普遍地受到关注,构建大数据时代下的国家网络安全已经上升到国家战略的高度。蒂姆·伯纳斯-李将下一代互联网称为“语义网(Semantic Web)”,并指出语义网就是“数据网(Web of Data)”,即数据是资源组织的单

位。因而,网络安全归根结底是数据的安全。在如今开放数据的浪潮中,数据安全更是制约开放数据的初衷能否最终实现的关键因素。

2010年,中国信息安全人才网(<http://www.cis-rc.com/>)成立,致力于搭建中国信息安全与保密专业人才的培养、实训和中外交流的高端平台。截至2012年11月,中国信息安全人才培养与就业工作研讨暨交流会已成功举办三届。中国信息安全博士论坛也已成功举办八届^[15]。北京小象科技CTO(首席技术官)洗茂源指出,通过培训奠定大数据基础技能的学员,在深入大数据安全,或者将自己以往的数据安全知识应用于新兴的大数据安全领域时,通过培训获得的知识具备更积极的作用^[16]。因此,有必要以信息安全人才网为平台依托,以上述两会为交流渠道,通过企业技能培训和学术交流等方式,为数据安全人才的培养和输出提供便利。

2.4 数据分析人才

这类人才主要从统计学、计算机科学、人工智能、可视化、信息经济学、网络科学、哲学社会科学、决策科学领域诞生,主要负责对大数据进行价值挖掘,包括对数据统计结果的甄别与分析,对数据分析结果的评估与展示,对用户数据需求的判断与反馈。南加州大学马歇尔商学院专门开设了商业数据分析的硕士项目,该项目介绍的第一句话就是:商业数据分析是现在全美增长最迅速的领域^[4]。

数据分析人才是最接近数据价值的一类人,现有的相似职位为数据分析师。百度百科中将数据分析师定义为在不同行业中,专门从事行业数据搜集、整理、分析,并依据数据做出行业研究、评估和预测的专业人员^[17]。也有说法是,数据分析师的主要职责包括寻找、检索、整理和传递从数据中来的见解,帮助报告和发现隐藏在数据潜在产品中的有意义的见解^[18]。笔者所指的数据分析人才是广义的,即在现有的数据分析师的基础上加上辅助决策。数据分析师是从大数据中提炼出有价值的见解,这一提炼过程的前提是准确和客观,即只要说出数据所隐含的事实就好,至于这个“事实”的价值判断则无关紧要。那么,把这个“事实”或“见解”摆在决策者面前,只有当决策者同样具备数据分析师的某些技能时,才可能做出“趋利”的决策,反之,则有可能在“事实”面前做出“趋害”的决策。我们知道,现在大部分企业的决策者并不是大数据时代的产物,因此他

们之中并没有太多人具备数据分析的能力。所以笔者提出数据分析人才, 他会在将“事实”和“见解”提炼出来的同时, 以仿真模拟的方式代替决策者做出决策, 至于决策者最终会不会接受这个决策就是另外一回事了。由于他亲历了数据分析的每一个步骤, 深知每一个因素对分析结果的影响, 那么, 在仿真模拟的时候, 便会在每一步都做出“趋利避害”的选择。同时, 他还具备全局观和系统观, 能够准确判断“1+1>2”的情形, 适时对偏差进行校正, 最后产生数据价值决策的最大化。

2.5 数据政策人才

这类人才主要从政策科学、公共政策学、公共管理学、社会学、伦理学、新闻传播学、法学、历史哲学、政治学领域诞生, 主要负责数据相关的政策、法律及制度的研究。

数据政策人才, 顾名思义, 是研究和制定有关数据政策的人才。大数据、开放数据都是社会公共领域的现象, 因此, 相关数据政策也属于公共政策的范围。对于公共生活涉及的诸多问题, 我们都希望“政策先行”, 然而在众多新运动、新革命中都是“实践先行”, 从技术革新中引发的大数据、开放数据运动就是其中的典型代表。即使是在西方崇尚法制的发达国家, 在这类运动中也是一次又一次徘徊、调整、修补相关政策法律文件。在实践中改进和完善, 本身是有益的, 只是在面临大数据、开放数据这类“双刃剑”运动的时候, 我们希望政策带有预见性, 成为走在前面为我们扫清阻碍的领路人, 而不仅是跟在后面为我们解决麻烦。

但在数据政策人才方面, 并不能像其他数据人才一样经过培训就大量产出, 在政策领域, 我们还是比较信赖“专家效应”的。那么, 数据政策人才的培养, 或许在起点上应该要稍高一些, 数据专家中研究政策的人员、政策专家中倡导数据运动的人员, 都可能成为今后所需的数据政策专门人才, 他们不仅肩负有关大数据、开放数据等数据运动政策研究的任务, 还担负着在具体政策制定和落实过程中为政府决策者提供专业意见的重任。

2.6 数据开放人才

这类人才的职责目前主要由倡导开放数据的各国政府首脑、互联网先驱以及数据公益组织首领承担, 随

着世界开放数据潮流的发展, 今后有望从统计学、信息技术、人工智能、网络科学、政策科学、社会学、经济学等领域诞生。他们主要负责开放数据的相关事宜, 如数据开放理念的传播和普及、开放数据运动的呼吁和推动、开放数据平台的建立和维护等。

开放数据是继自由软件、开源运动和开放存取后又一崇尚开放、自由、共享精神的热点。开放数据协会(ODI)的首席执行官加文·斯塔克斯(Gavin Starks)曾说, “开放数据像极了1994年时的互联网, 正如网络出现早期我们所看到的潜在发展机会那样, 开放数据的潜力同样不可小觑”^[19], 这个潜力的挖掘, 就需要依赖数据开放人才的发现。数据开放人才并不一定要具备深厚的技术功底, 他所应具备的素质和能力, 首先是对“开放”理念的认知与认可, 其次是推动“开放”的理想, 最后是具有一定影响力的开放数据实践。Julie McLeod在探讨开放获取和开放数据对信息和记录管理者产生的潜在挑战和机遇时表示, 在新的环境背景下, 他们专业知识的应用受到来自技术和竞争对手的挑战^[20]。所以, 如果除了上述三点之外, 还具备一些数据挖掘、数据建模、数据分析等技能的话, 开放知识的应用就会相对简单, 也能大大缩短数据开放人才培养和产出的周期。

2.7 数据科学家

笔者检索研究发现, 绝大多数描述“数据科学家”这一新兴职业的文献都强调, 数据科学家需要独特的综合技能, 在数据方面是典型的“十”字型人才, 即至少拥有两项以上上述的数据技术、数据管理、数据安全、数据分析、数据政策、数据开放人才所必备的基本技能。

美国商业分析软件与服务供应商SAS公司大中华区总裁吴辅世指出: 在大数据时代, 数据科学家等分析人才的需求将激增, 尽早开始人才储备将是企业稳步发展的优势之一^[4]。网址缩短服务公司Bitly的首席科学家Hilary Manson曾定义大数据科学家为: 能获取、清洗、探究、建模和诠释数据的人^[21]。人人游戏高级数据科学家陈弢、百度大数据首席架构师林仕鼎, 都将数据科学家分为广义和狭义两种: 从广义的角度, 以数据为处理对象的从业者都可称为数据科学家, 比如原来的数据库管理人员、数据架构师、数据库工程师和数据统计分析师; 从狭义的角度, 只有那些能够利用数据

作为资源,具有数据分析能力,精通各类算法,直接处理数据,创造附加价值的人员才可以称为数据科学家^[22-23]。

结合陈弢、林仕鼎、Talent Analytics公司首席执行官Greta Roberts、中桥调研咨询首席分析师王丛、领英网首席数据科学家Manu Sharma等人对数据科学家应具备素质和能力的探讨,笔者认为数据科学家应具备以下四种基本素质和能力。首先,应具备遵守伦理道德、政策法律的基本公民素质,即在数据处理、管理和应用过程中,保护公民的隐私权、数据权等;其次,应具备科学家的基本素质,即客观、诚实、严谨、创新、坚韧和好奇;再次,应具备对于数据的处理能力,即掌握编程(存在争议)、算法、数据采集、数据统计、数据整理、数据建模、数据挖掘、数据分析、数据可视化等方面的能力;最后,应具备涉众的业务能力,即市场调研、数据应用、交流沟通(企业内外部)、业务开发、业务管理、业务服务、辅助决策、挖掘数据的商业价值等方面的能力。

王丛认为,数据科学家的知识储备比例应该是:业务管理流程经验占40%,数据分析流程和与数据相关的基本理论知识占30%,支持大数据分析的IT架构方面的知识占30%^[22]。不难看出,数据科学家是所有数据人才中“最需要动脑筋的人”,要成为这样的人才,无论是知识储备,还是专业技能,都受到很大的挑战。可以通过从前面六种数据人才中直接择优培养的方式来缩短数据科学的培养周期。

3 大数据时代我国数据人才的培养路径探析

数据人才已经成为当下最热门的高素质人才缺口,随着大数据以及开放数据进程的加快,全球范围内的数据人才争夺战愈演愈烈。冼茂源在谈到解决贵州大数据人才短缺问题时指出,解决大数据人才短缺,短期可以靠引进,长期还是要靠本地培养^[16]。因此,以本地培养和产出为最终目标,笔者从以下三个方面探讨我国数据人才的培养路径。

3.1 积极开展政策研究,建立政策扶植体系

作为国内的大数据领跑省份,贵州省近年来对数据人才的需求量越来越大,为解决人才短缺的现状,

政府在政策上予以了大力支持,2014年发布的《关于加快大数据产业发展应用若干政策的意见》和《贵州省大数据产业发展应用规划纲要(2014-2020年)》提出对于大数据产业人才,将积极纳入省“百千万人才引进计划”,充分享受优惠奖励政策^[24]。当然,这些政策主要是为了吸引和鼓励大数据人才在贵州能进得来、留得住,但依据贵州的政策思路,国内其他省份也可以效仿。对于“引进”的数据人才,给予户籍、医疗、子女上学、住房补贴、个税返还奖励等多方面的优惠政策。值得注意的是,虽然受到地理位置、气候环境、经济发展及生活交通等多方因素的影响,可能同样的政策扶植会导致数据人才的“趋多避少”,而造成人才分布的严重不均。但从长远来看,人才之间的竞争对我国总体数据人才质量的提高是有益无害的,并且这个问题也可以通过区域特色政策和差别性政策得到缓解。人才的引进只能解决短期问题,在通过政策手段让“引进”人才本地化的同时,也应建立政策扶植体系,培养本地数据人才。例如鼓励引进的数据人才开办培训班,吸引有兴趣的本地IT人员加入学习;资助本地IT精英外出深造,并且通过一系列的优惠政策确保其学成之后能够回馈本土。

以上是对各省级行政区域的政策建议,在国家层面,国务院2010年印发的《国家中长期人才发展规划纲要(2010-2020年)》已将人才培养提升到国家战略的高度,而数据人才作为近年来经济社会发展重点领域急需紧缺的专门人才,理应成为《纲要》的首要发展对象。笔者建议,以《纲要》的任务为根本,研究和颁布专门的数据人才培养政策或意见,全面协调现有数据人才在我国分布不均的现状,改革传统的人才培养模式,对数据人才实行分级培养和产出,在确保人才质量的同时尽量缩短培养周期。

3.2 改革高校人才培育体制,将数据人才培养纳入国家高等教育体系

高等院校作为我国人才培养的主要基地,承担着大部分高素质人才产出的任务,在大数据时代也不例外。面对数据人才严重缺乏的现状,目前国内外高校都采取了积极的应对措施。在国外,美国纽约大学、英国邓迪大学均从2013年起设立数据科学硕士学位,美国南加州大学、纽约大学已开设数据科学专业(如南加州大学马歇尔商学院专门开设了商业数据分析的硕士项

目),成为2015年中国留学生热门专业。同时,美国哥伦比亚大学从2015年起设立数据科学博士学位。此外,美国有20所顶尖高校开设了“大数据硕士学位”项目,包括哈佛大学、斯坦福大学、加州大学伯克利分校、弗吉尼亚大学、哥伦比亚大学、俄亥俄州立大学等^[4,6,25]。

在国内,香港中文大学自2008年起设立了“数据科学商业统计”科学硕士学位。复旦大学2013年成立数据科学研究中心。2014年4月26日,清华大学与青岛市人民政府签署共建合作协议,正式成立清华-青岛数据科学研究院,并宣布将推出多学科交叉培养的大数据硕士项目,9月开始培养第一批大数据硕士学位研究生。2014年7月26日,北京大学成立北京大学大数据技术研究院,学校的信息管理系统也特设了情报学专业(大数据方向)的管理学硕士高级研修项目。中国科学院也从2014年开始招收大数据技术与应用方向计算机技术工程硕士研究生。此外,西安交通大学、浙江大学、华东师大等高校也先后设立了数据科学研究中心^[25-27]。2014年6月,上海市科委启动了“数据科学和大数据人才培养计划”,将在未来3年培养和引进千名高端数据人才。该培养计划首批启动了大数据工程硕士、数据科学家训练营和数据科学FIST三个项目,随后还将陆续启动青年数据科学家培养、数据科学博士学位、数据科学硕士学位、数据科学专业硕士学位和数据科学本科第二专业学位等6个培养项目^[28]。

值得一提的是,在高等院校之外,国内部分中等学校也开始进行适时的课改。2010年9月,在北京师范大学项华主持的北京市教育科学“十一五”规划课题“数字科学家计划——基于数据探究的选修课程设计与试验”的指导下,北京景山学校设计了“数字科学家”课程(属于信息技术课程)。现在,“数字科学家”已经成为该校的特色校本课程,能够有效培养中学生的数据探究意识和数据探究能力^[29]。

国内外高校和景山中学的课改给予我们很大的启发。诚然,大数据人才的直接输出大多由高等院校完成,但是站在人才基础素质培养和可持续发展的角度,中等学校通过信息技术课程培养学生的数据意识和能力是值得肯定与鼓励的。这也为我国其他类型人才的培养提供了启示,即一名优秀的专业人才可以通过中等学校兴趣培养、高等学院技能培育、社会实践能力锻炼,最后作为相应人才供给输出。当然,目前需要大量的数据人才以解燃眉之急,切实有效的方法是从技能培育阶段的高校入手,通过改革传统的人才培养体制

和学科教育体系,将“数据”纳入教育范畴,参考国内外高校设置“数据科学”学士、硕士、博士学位,并积极建立高校主导的数据科学研究中心,为数据人才的输出做好知识和技能储备。

3.3 加强产学研结合,开展数据人才培养实训基地的建设

有专家指出,由于无法提供真实的大数据环境,高校很难培养出市场真正需要的大数据人才,于是“产学研”相结合的实践陆续推出。

近年来,贵阳正在实施大数据人才“贵阳造”计划,即通过政校企合作的方式,为贵阳培养大数据基础性实用技术人才。2013年9月底,国内顶尖的云计算研发和营运公司北京讯鸟软件有限公司落户贵阳南明区,成立贵阳讯鸟云计算科技有限公司,仅20天后,贵阳讯鸟便在贵阳市政府的协助下与贵州财经大学签订校企合作协议,双方联合成立云计算研究实验室、人才培养基地。2014年5月,大数据行业巨头甲骨文软件系统公司也瞄准了贵阳的大数据人才培训市场,计划以校企合作、开设专业培训基地等方式,抢占市场先机^[9]。2015年年初,国内知名大数据在线教育平台小象学院与贵州大学、贵州财经大学等贵州高校商谈合作开设大数据专业,双方分担师资力量试点开班,吸收来自计算机、统计学乃至视觉设计等不同专业背景的学生进行大数据人才培养^[16]。2015年8月21日,致力于在中国产学研结合、打造开放社区式创新生态系统的ARM宣布与贵州大学合作建立“贵州大学—ARM创新与人才培养基地”,旨在推动当地大数据教育发展,为打造可持续发展的“云上贵州”培养大数据人才^[30]。

此外,2015年3月,IBM与香港中文大学市场学系、对外贸易大学国际商学院、西南交大经济管理学院等联合宣布推出“百企大数据A100”计划。加入该联盟的高校将向100家拥有B2C数据的企业投放专业的教授、研究生及本科生,帮助企业进行数据库整合、数据库挖掘、市场决策支持、产品推荐、社交聆听等大数据领域的分析和研究^[31]。

由于以上数据人才培养实训基地落成的时间都不长,还未能有定量分析数据来展示这些项目的成果,所以在此不做绝对的判断。只是实训基地的建成,为高校学生提供了学校没有的数据环境和实战机会,就人才培养过程来看,是有益的。因此,就培养数据人才

业务应用能力和综合实践能力来说,采取“产学研”结合,积极开展数据人才培养实训基地建设的方法较为可行。

4 图书情报学科领域在数据人才培养中的定位与前瞻

大数据的兴起,首先是在信息技术领域,IT技术的革新,使数据从“量”变到“质”变而形成大数据,进而在相关领域引发大追逐和大讨论。这其中又以与信息科学、计算机科学联系紧密的图书情报学科为代表,数字图书馆、档案数字化、移动图书馆、移动阅读、情报学专业的大数据研究方向等,都是大数据时代图书情报领域的产物。在人才培养方面,以黑龙江大学信息管理学院为例,在本科设置的信息管理与信息系统和电子商务专业中,都开设了数据相关课程,如C程序设计、C++、C#.NET、VB程序设计、数据结构、Java程序设计、数据库、计算机网络、网络编程、软件测试、系统工程等;在硕士研究生设置的情报学、图书馆学专业中,开设了信息资源管理、决策支持系统、网络信息组织与检索、数字参考咨询、信息分析与预测等专业课程。不难看出,一些课程的设置(C程序设计、网络编程、数据库等),有助于培养学生的数据处理技术能力,这些能力也是当今数据科学家所必备的。还有一些课程(信息资源管理、数字参考咨询、信息分析与预测等),对于培养学生的数据意识和数据思维有一定的帮助。此外,学院还于2003年建立了“信息管理学院创业教育基地”,旨在培养学生的学习能力、实践能力和团队协作能力,对计算机编程、网页设计等相关内容感兴趣的学生都可以参加。近年来,从基地走出的应届毕业生就业竞争力普遍比一般学生要高一些,就职单位大多在知名的互联网大数据企业,如腾讯、百度、阿里巴巴、新浪、搜狐、完美世界等。以黑龙江大学信息管理学院为缩影,全国其他高校的图书情报教学单位对数据人才的培养也都有一定的考量,并且在如今数据人才紧缺的情况下,也必将加大培养力度。

但不可否认的是,图书情报领域对于数据人才的培养是有限的。它既不能像计算机科学、软件工程学那样,培养出IT精英人才,也不能像统计学那样培养出数据“敏感”型人才。虽然针对大数据时代所需数据人才的综合能力,在图书情报学科建设者的努力之下,培养

出的人才兼而有之,但由于数据和大数据的特殊属性,对数据人才必备能力的要求甚为严格,图书情报现有培养体制培育出的人才与现实需求尚有一定的距离。因此,如何发挥图书情报、信息管理学科的传统优势,结合社会对数据人才的现实而迫切的需求,改变教育理念、改革培养体制、变革培养手段,积极融入大数据时代的数据人才培养体系并形成自身特色,是值得我们深入思考、系统重构的涉及教育改革与学科发展的重要问题。我们提出的基本思路包括:①数据人才培养是一项新的跨学科跨领域的宏大社会系统工程,图书情报学科应当坚持有所为、有所不为的原则,在积极融入新的数据人才培养体系中把握机遇,充分认识自身的优势、合理规划自身的定位,以大数据时代的社会需求改造图书情报学科人才培养体系及课程设置,寻找本学科专业新的增长点;②科学规划并构建基于图书情报学科的数据人才培养机制、体现本学科专业的特色与时代价值,即以数据驱动为背景,以数据共享为目标,以数据管理人才类型为定位,以跨界复合为模式,以数据挖掘、数据分析等专业技能为核心,以开放数据能力、数据评价能力、数据统计能力等职业能力为重点,抢抓机遇、拓展生存空间,打造新型人才培养方案及课程设置体系,为大数据社会培养专门化适格人才;③鉴于目前设立数据管理类专业的条件尚不成熟,借鉴上文3.2中分析的国内外高等学校设立数据类人才培养方向的经验,图书情报学科可以考虑在研究生培养层面、尤其是面向管理和实践的图书情报专业硕士培养机制中设立数据管理类研究与培养方向,利用图书情报专硕培养机制相对灵活的优势,尽快介入数据类人才培养的教育体系中并展示图书情报学科的特殊贡献。

5 结语

数据人才短缺的问题短期内无法得到彻底解决,而目前对数据人才的多种培养措施也都在摸索阶段。作为人才培育的主要阵地,高校应该义不容辞地承担起这份艰巨的任务,适时改革人才培育体系,积极探索校企联合新模式,从而不仅做到堵住现有的人才缺口,还能实现数据人才的可持续发展。图书情报学科领域也应该重视自己在这项任务中所担任的角色,正视自身的优势与不足,及时扬长避短,以在数据人才培养这个新时期新任务中有所作为。

参考文献

- [1] 卢朵宝. 美国学者质疑“大数据”理论[N]. 经济参考报, 2013-06-14(008).
- [2] Simón L F R, Avilés R A, Botezan I, et al. Open Data as Universal Service. New perspectives in the Information Profession [J]. Procedia - Social and Behavioral Sciences, 2014, 147: 126-132.
- [3] 2015第一届数据科学家大会[EB/OL]. [2015-12-23]. <http://datascientist.fudan.edu.cn/>.
- [4] 吴成良. “数据分析”成了“金饭碗”[N]. 人民日报, 2014-08-21(022).
- [5] 涂恬. “数据科学家”正在成为最热门职业[N]. 科技日报, 2014-06-13(005).
- [6] 老鬼阿定. 未来属于数据科学家[N]. 计算机世界, 2014-05-26(008).
- [7] 数据科学家到底有多稀缺[J]. 商学院, 2015(8):14-15.
- [8] 陈桂龙. 九三学社中央: 加强大数据人才队伍建设[J]. 中国建设信息, 2015(7):21.
- [9] 谢江林. 国内外巨头看好贵阳市场前景, 入筑分食大数据人才培养蛋糕[N]. 贵阳日报, 2014-05-27(006).
- [10] 国家中长期人才发展规划纲要(2010-2020年)发布[EB/OL]. [2015-12-25]. http://www.gov.cn/jrzq/2010-06/06/content_1621777.htm.
- [11] 董晓婷. 大数据的定义特征及其应用分析[J]. 硅谷, 2013(11):120.
- [12] 程赏. 数据库技术仍在演进, 关系型数据库难以被替代[EB/OL]. [2015-12-25]. http://tech.ifeng.com/internet/detail_2014_07/11/37295276_0.shtml.
- [13] 数据管理[EB/OL]. [2015-12-25]. <http://baike.baidu.com/view/14717.htm>.
- [14] 2015中国网络安全大会[EB/OL]. [2015-12-12]. <http://nsc.skdlabs.com/>.
- [15] 第八届信息安全博士论坛[EB/OL]. [2015-12-25]. <http://www.secdocor.com/plus/list.php?tid=828>.
- [16] 鲍夏挺. 解决贵州大数据人才短缺问题短期可以引进长期要靠高校[N]. 贵阳日报, 2015-02-25(005).
- [17] 数据分析师[EB/OL]. [2015-12-25]. http://baike.baidu.com/link?url=kaYtTnmEbO4Amdw6DITWaKShGxSEWZ53TNIONLZaUNMracNAbebdk2ExQlYonrQ2n_-tPILTnb6tnoiZAKR6q.
- [18] HanKai. 成为一名数据分析师的新手指导[EB/OL]. [2015-12-25]. <http://blog.jobbole.com/84606/>.
- [19] 乔尔·古林. 开放数据[M]. 张尚轩译. 北京: 中信出版社, 2015:4.
- [20] McLeod J. Thoughts on the opportunities for records professionals of the open access, open data agenda [J]. Records Management Journal, 2012, 22(2): 92-97.
- [21] 沈建苗. 如何成为大数据科学家[N]. 计算机世界, 2013-07-08(T08).
- [22] 郭涛. 揭秘数据科学家[N]. 中国计算机报, 2013-10-21(008).
- [23] 企业需要三类大数据人才[N]. 中国计算机报, 2013-10-21(010).
- [24] 杨春凌. 贵州广揽大数据人才[N]. 贵州日报, 2014-03-30(001).
- [25] 沈淑莎. 上海启动大数据人才培养计划[J]. 现代人才, 2014(3):8.
- [26] 刘蔚如, 林笼. 清华成立数据科学研究院[N]. 新清华, 2014-05-09(004).
- [27] 徐扬. 新时期大数据人才的培养[J]. 中国劳动, 2015(17):28-29.
- [28] 文汇报. 上海启动大数据人才培养计划 首批工程硕士招生6月报名[EB/OL]. [20133403586.htm].
- [29] 毛澄洁. 让学生与技术发展同步, 做大数据时代的数字科学家: “数字科学家”课程的开发与尝试[J]. 中国信息技术教育, 2013(11):94-97.
- [30] ARM携手贵州大学共建大数据创新与人才培养基地[J]. 微型机与应用, 2015(17):74.
- [31] IBM携手多家高校推出“百企大数据A100”计划[EB/OL]. [2015-12-25]. <http://www.cit360.com.cn/News/ITxw/7685.html>.

作者简介

马海群, 男, 1964年生, 博士, 黑龙江大学信息管理学院教授、博士生导师, 研究方向: 信息政策与法律, E-mail: mahaiqun@sina.com.cn。
蒲攀, 女, 1990年生, 黑龙江大学信息管理学院在读研究生, 研究方向: 信息资源管理。

The Ponder of Cultivation for Talents in Data under the Circumstance of Big Data in China

MA HaiQun¹, PU Pan²

(1. Research Center of Information Resources Management, Heilongjiang University, Harbin 150080, China;
2. College of Information Management, Heilongjiang University, Harbin 150080, China)

Abstract: The analysis is about the current status of the requirements of talents in data under the circumstance of big data at home and abroad, and this paper divides the talents in data into seven types: talents in data technology, talents in data management, talents in data security, talents in data analysis, talents in data policy, talents in data opening and talents in data scientist. On the basis of deep analysis, three ways are proposed for the cultivation of talents in data of our country, and then we point out the location and prospective of the discipline of library and information science for the cultivation of talents in data in China.

Keywords: Talent in Data; Personnel Training; Discipline of Library and Information Science

(收稿日期: 2015-12-25)