

【科研管理】

# 大数据环境下的高校科研管理信息化探索<sup>\*</sup>

许哲军 付 尧

(华东理工大学 科学技术发展研究院,上海 徐汇 200237)

**摘 要:** 近年来,大数据引起了产业界、学术界以及政府部门的高度关注。高校作为我国科研创新的重要基地,科研信息化水平极大影响到高校创新事业的发展。目前我国高校科技管理信息化存在数据共享度比较低,科技产出数据零散不系统,数据的质量不理想,信息化数据的利用率低等问题。论文简要阐述了大数据的概念以及传统模式下的科研管理的挑战,并提出了基于大数据技术的科研管理信息化集成解决方案。

**关键词:** 大数据; 科研管理信息化; 数据挖掘

中图分类号: G 526.3

文献标识码: A

文章编号: 1672-7312(2014)02-0112-04

## Exploration of Scientific Research Management Informationization in Context of Big Data

XU Zhe-jun, FU Yao

(Institute of Science and Technology Development, East China University of Science and Technology, Shanghai 200237, China)

**Abstract:** In recent years, big data caused more attention in industry, technology and government. In our country, colleges and universities are important bases of scientific research and innovation. The development of national science and technology is greatly affected by the information service level of scientific and technical research management in colleges and universities. At present, there are many problems in scientific research management informationization, such as low level of data sharing, low quality and low utilization of data, fragmented and non-systematic data of scientific output. This paper briefly described the concept of big data and the challenges to the traditional scientific research management, then suggested the application of big data technologies in scientific research management informationization.

**Key words:** big data; scientific research management informationization; data mining

### 1 引 言

近年来,大数据概念引起了科技界、产业界和政府部门的高度关注。Nature 和 Science 等国际顶级学术刊物相继出版专刊来专门探讨对大数据的研究<sup>[1]</sup>,从多个方面介绍了海量数据带来的挑战,特别指出,倘若能够更有效地组织和利用这些数据,人们将得到更多的机会发挥科学技术对社会发展的巨大推动作用;大数据的开发与利用已经在医疗服务、零售业、金融业、制造业、物流、电信等行业广泛展开,并产生了巨大的社会价值和产业空间<sup>[2-3]</sup>;政府

部门也高度重视大数据技术,2012年3月奥巴马宣布美国政府投资2亿美元启动“大数据研究和发展计划(Big Data Research and Development Initiative)”。2013年初,中科院倡议应将大数据提升为国家战略。由此可见,大数据确实引起了热议并形成共识。

自上世纪末推行高校信息化以来,很多高校都建立和运行着各类不同的数据库系统,如学生系统、教务系统、科研系统等等。但是,存在于各部门之间的数据库系统并不关联,由此,在全校范围内形成了多个信息孤岛,不但导致了校内大量的资源、资金浪

\* 收稿日期: 2013-12-12

作者简介: 许哲军(1978-),男,江苏宜兴人,博士,主要从事数据挖掘、机器学习、科研管理方向研究。

费严重,还给教学、科研的深入管理带来了困难。随着科研决策系统对各种数据的依赖程度的不断加深,这就需要综合利用和分析各个部门的数据。如何有效的集成、深入分析这些数据已成为科研管理进一步深化所急需解决的问题。需要有新的更有效的技术对大量数据进行分析以发挥其潜能,从中迅速提取有用的信息以指导和辅助科研管理。因此有必要在科研管理中引入大数据技术,以提高高校科研管理的水平和能力。

## 2 大数据简介

所谓大数据,就是用现有的一般技术难以管理的大量数据的集合。大数据的特征,通常用四个V开头的关键词来描述,一是Volume(容量),也就是数据体量大,已经从TB级别跃升至PB、EB级别(1 024 GB = 1 TB, 1 024 TB = 1 PB, 1 024 PB = 1 EB, 1 024 EB = 1 ZB),根据IDC的一份名为“数字宇宙”的报告,预计到2020年全球数据使用量将会达到35.2 ZB;二是Variety(多样性),即数据类型繁多,数据来自多种数据源,数据种类和格式日渐丰富,已经冲破了以前所限定的结构化数据范畴,囊括了半结构化和非结构化数据。三是Velocity(速度),数据产生和更新的频率,也是大数据的一个重要特征,全球数据量每18个月翻一番(遵循摩尔定律)。四是Value(价值),数据价值密度低,海量的数据可能包含极少量的有价值的信息,如何高效提取这些有价值的信息是关键。

从各种各样类型的数据中,快速获得有价值信息的能力就是大数据技术。大数据技术是解决数据丰富而知识贫乏的有效途径,其实质是从数据中提取隐含的、未知的和潜在有用信息的过程,被公认为是数据库研究中的一个极富应用前景的新领域。掌握这一技术已经成为一种新的竞争优势。对于政府,是构建高效服务型政府的关键;对于高校,大数据为高校绩效管理、科研评估、研发资源分配等提供了有效的决策支撑。

## 3 传统高校科研管理中引入大数据技术的必要性

### 3.1 传统高校科研管理信息化存在的问题

随着211院校建设、高校信息化的不断推进,在高校运行着的各种系统和各类数据库,如教务系统、学工系统、研究生系统、人事系统以及运行于科研管

理部门的项目管理系统、经费管理系统、成果管理系统等。尤其是科研管理信息化的推进简化科技处的日常数据管理工作,极大地提高科技管理的工作效率,积累了大量的数据。但在当前国家创新大背景下,国家科技投入不断加大,高校科技活动迅猛发展,对科技管理信息化的要求越来越高,而现行的科研管理技术也存在一些不足之处:

首先,管理信息化目前仅停留于数据收集阶段。管理系统的功能主要集中于项目、经费、成果的录入、修改、查询、报表等基本功能,管理人员只能通过简单的统计或排序等功能获得表面的信息,隐藏在这些大量数据中的信息一直没有得到有效的应用<sup>[4]</sup>。

其次,统计分析功能比较简单,与高校其他数字化平台的关联度低。系统的统计分析功能绝大多数局限于对人员、科研经费、论文数、专利等的简单累加,很少给出数据反映的问题。诸如教学管理系统、人事管理系统、科研管理系统等等,这些系统之间基本上没有实现信息的共享。

第三,数据积累多,辅助决策功能欠缺。随着各类系统的不断使用积累了大量的原始数据,但这些数据背后隐藏着什么样的信息?能否通过对数据的多角度分析为相关人员提供更加丰富和有利的决策支持?目前的科研管理系统的数据分析功能还是很有限的。

因此,通过将大数据的概念以及技术引入到传统的科研管理中,对积累的数据进行重新利用,将现有的管理数据转化为可供使用的知识,以此提高科技管理水平和技术含量,为管理部门决策提供客观、科学、全面的参考。

### 3.2 科研管理中引入大数据方法的优越性

大数据技术是网络技术、人工智能、数据库技术等现代信息技术的有效结合。具有无可比拟的先进性;大数据技术通过针对科技管理系统、人事系统、教务系统以及基于互联网的大型科技文献数据库、专利数据库等信息资源的关联分析,找出数据的相关性,提取有价值的信息,可以为传统的专家定性决策管理提供广泛、深入的数据支持。同时,大数据技术是从大量数据中发现那些尚未发现的知识,是从科学技术活动大量原始数据中自动获得知识和重要信息的过程<sup>[5]</sup>。这些知识直接来源于数据库内部,因此它不受外部资源的限制,具有相对的独立性。

大数据技术让数据产生知识,将数据管理工作提升为分析预测<sup>[6-7]</sup>。数据衍生知识的过程如图1

所示,首先是通过收集来自高校内部和外部的各种不同的数据源,提取有用的信息,经过数据清洗、转换、重构进入数据仓库,然后通过合适的查询和分析工具、数据挖掘工具、联机处理工具对数据进行加工处理,最终转化为知识。图中内部源数据包括校内项目、经费、人员、设备等数据库信息,外部源包括主管部门、各地省市自治区科研数据;WEB 数据源为各大公开的文献、专利、成果等数据库。

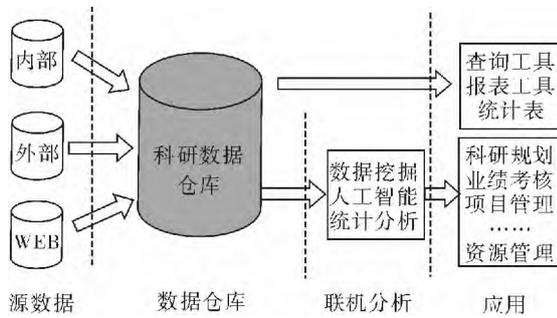


图 1 数据衍生知识流程示意图

#### 4 大数据方法在科技管理中的应用探索

如何建立科学、高效的科研管理模式,优化科研资源配置和提高科研计划项目的管理绩效,保障高校科研事业的良性、健康发展,已经成为当前高校科研管理体制改革的的一项重要内容,大数据方法在这一领域的实施具有广泛的应用前景。

##### 4.1 科研评估

科研评估是现代科研管理的一种重要手段,科研资源的合理有效分配、利用和管理以及科研项目和科研机构全面客观的评价等各个方面都对科学研究的绩效评估提出了新的要求。

在科研评估体系中,数据的处理在其中占有非常重要的作用,随着高校信息化的推进,以及大型科技文献、论文、专利数据库系统等在科学研究中的普及,全球范围内的科技数据信息量急剧增大。面对庞大的科技信息数据库,需要快捷地从中提取出有用和有效的知识,因此巨大的科技信息数据有待分析处理。面对这种海量数据的增加,多数科研部门仅使用传统的数据分析和简单的传统统计方法进行数据分析,由于受到人力、物力、财力的限制,数据背后隐藏的深层次知识无法有效得以理解使用,相反却带来了“数据灾难”和“数据荒废”。用传统的统计分析手段已很难以满足要求,若需要发现更深层次的规律,并保证系统运行的经济性、安全性和可靠性,还必须提供更快、更有效的决策支持。而大数据

技术恰好满足了这一要求。

大数据技术可以综合内部、外部以及 WEB 数据资料:从主管部门管理系统中获得研究项目的数量与类别,从内部数据库得到人员、经费、设备等信息,从 WEB 上获得论文和专利的数量和质量,从项目成果报表得到获奖、成果转让情况,最终综合集成各类数据。通过建立综合评估模型将各类指标进行整合,为科研评估的专家评议提供数据支持,最后得出评估结果。

##### 4.2 项目立项决策

项目立项决策是在科技活动实施前对实施该项活动的必要性、可行性及其定位、目标、任务、投入、组织管理等所进行的评价,主要是为项目立项决策提供依据。科技项目立项评价内容主要包括三个方面:一是项目的科学性,包括立项依据、科研内容创新性、项目的效益等;二是项目方案的合理性,即经费安排是否合理、技术路线是否可行;三是项目的运行条件评价,包括项目申请人的研究能力、工作基础以及前期项目的完成情况。项目立项评价的主要目标是对项目的优劣进行判断,是科研管理的重要内容之一。

国内科研项目立项主要采取由科研人员申请,科技主管部门组织专家评审论证筛选,然后择优选择承担单位和负责人的管理模式。科研立项涉及到大量的数据管理量,包括课题申请单位、申请人信息、课题信息、经费信息、评审专家信息等方面数据。目前虽然各级科研管理部门的数据库管理系统基本涵盖了项目申请和评审过程中的大部分日常操作,如项目申报上传、分发给评审专家、自动统计评审结果等一系列项目申报中的常规性事务性工作,但这些功能只是在传统管理模式上的一个信息化过程,基本没有任何的辅助决策功能。

如何利用已有的内部以及外部数据,进行数据分析和挖掘工作,减少科研立项工作中存在的一些重复性、经费安排不合理、项目依托人不称职等因素,以此指导科研项目立项,促进科技资源优化配置,提高科技经费的使用效益,促进公平竞争。这项工作本身是一项重大的研究课题,由于大数据概念的引入,可以通过大数据技术,首先对课题的科学性、创新性与外部文献库进行结合分析;其次申请者所涉及的各项因素进行多数据的联合查询和分析,发现和建立科学的指标体系和项目筛选方法,将不合理因素排除于项目立项之前,合理地选择项目予以资助。

### 4.3 优化科研资源配置

我国的高校一方面存在着科研资源短缺的状况,另一方面又存在着科研资源未被很好地利用,形成很大浪费的现象。同时还存在着科技资源配置的不同院系之间的分割明显,不同学科间的以及产学研之间的分隔,这些制约了全校整体科技创新活动的统筹协调,造成了严重的资源浪费和低水平重复建设,制约着科研管理效率,限制了高校的整体创新和科研实力。

基于大数据技术的科研资源配置模式可分为三个层面,首先是数据采集和清洗,建立包括人员库、成果库、经费库以及以科研用房、科研仪器为基础的科研资源数据库;其次是建立适合各自高校学科发展的评判模型库,包括各类科研资源、科研成果的计算参数和规则库;第三是以定量化绩效考核为基础的资源配置工具和决策支持管理工具,以此完成基于大数据技术的科研资源优化配置过程。

### 4.4 科研项目的全过程管理

传统的科研管理“重申请、轻结题”,忽视项目的过程管理,导致项目的完成质量完全依赖于项目负责人的责任与态度。申请与结题管理指标明显,提高项目中标率和成果获奖率往往是科研管理人员追求的目标,效益明显,管理程序已固定;而中期管理伸缩性大,效益潜在,管理较复杂,是项目管理中的难点。

科研项目的全过程管理,是指项目立项后到结题前的全过程的监管<sup>[8-9]</sup>。管理内容包括科研项目计划的制定、人员管理、财务管理、中期检查等等。科研项目的全过程管理可以从信息化工作入手,对分散的各类相关数据库的信息进行综合分析和提炼,跟踪财务系统的项目经费使用信息、科技文献数据库的项目相关成果信息以及引用情况、课题组成员的教学工作量以及仪器设备的使用情况等,对这些数据的联机和数据挖掘可以发现课题承担人、承担单位、经费分配、科研仪器使用等多种相关信息之间的内在关系,从而提高科研管理者及时发现问题、解决问题的能力。

### 4.5 科研规划

高校科学研究能力是衡量一个国家基础研究和高技术前沿领域原始性创新能力的重要标志,是国家发展创新科技的先锋,对各国未来能否在日趋激烈的全球科技竞争中占据有利地位具有举足轻重的影响。从高校自身发展来看,科学研究是学科建设

的重要承载和巨大推动,对高校发展具有深远的影响。在快速多变、竞争加剧、形势日趋复杂的背景下,机遇和挑战与日俱增,高校科研战略规划的重要性日益显著。

通过对大型数据仓库进行有效的挖掘,可以对一个单位所关注的关键技术,重点领域和发展方向进行分类和预测。通过建立模型、数据可视化和生成文本报告等形式向管理者提供各种影响因素之间的内在关联,以指导科技发展规划的制订。

## 5 结语

大数据及其数据分析技术是一个多学科交叉融合而形成的新的研究热点,已经被广泛应用于市场分析、电子商务、政府部门、金融业和知识管理等方面,在多个领域取得令人满意的应用效果。在科技管理领域里,随着信息化的不断推进,数据积累量的不断增长,把大数据技术应用到高校管理领域中,可以进一步促进管理体制的改革,完善和发展在管理与决策中凭经验或学习别人经验的做法,同时,大数据方法能较客观地反映高校科研管理中存在的问题,为决策提供重要数据支撑。

### 参考文献:

- [1] Goldston D. Big data: Data wrangling [J]. Nature, 2008, 455(7209): 15-15.
- [2] Bughin J, Chui M, Manyika J. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch [J]. McKinsey Quarterly, 2010(8): 1-14.
- [3] Gantz J, Reinsel D. 2011 Digital universe study: extracting value from chaos [M]. IDC Go-to-Market Services, 2011.
- [4] 吴生, 赵雪曼. 高校科技统计实践与分析 [J]. 技术与创新管理, 2012(33): 503-505.
- [5] 张高亮. 基于知识发现的科研管理系统模型 [J]. 西南师范大学学报: 自然科学版, 2001, 26: 393-396.
- [6] 杨国梁, 关忠诚, 石兵. 数据仓库技术在中国科学院评估数据分析中的应用 [J]. 科研管理, 2007, 14: 114-122.
- [7] 郭卜铭, 吕渭济. 高校科研管理中的数据挖掘技术及应用 [J]. 科技和产业, 2007(7): 38-40.
- [8] 宋永杰. 科研经费全过程管理的探讨 [J]. 中国科技论坛, 2009(11): 3-7.
- [9] 许长谭. 高校科研项目全过程管理模式研究 [J]. 四川文理学院学报: 社会科学版, 2008(18): 87-89.